UCLA
Experiments in Linguistic Meaning

Title
Tracking the activation of scalar alternatives with semantic priming

Permalink
https://escholarship.org/uc/item/6jx8c5k8

Journal
Experiments in Linguistic Meaning, 2(1)

Authors
Ronai, Eszter
Xiang, Ming

Publication Date
2023-01-27

DOI
10.3765/elm.2.5371

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Tracking the activation of scalar alternatives with semantic priming

Eszter Ronai & Ming Xiang*

Abstract. From an utterance of *Mary ate some of the deep dish*, hearers frequently infer that Mary didn't eat all of the deep dish. Similarly, an utterance of *The movie is good* might lead hearers to conclude that the movie isn't excellent. These inferences are instances of scalar implicature (SI). The standard assumption is that SI arises via hearers' reasoning about alternative utterances that the speaker could have said, but did not. In particular, hearers are taken to consider stronger alternatives such as *all* (or *Mary ate all of the deep dish*) and *excellent* (or *The movie is excellent*) and derive their negation. In this study, we investigate the psycholinguistic reflexes of this inferential process. We use semantic priming with lexical decision to test whether lexical alternatives such as *all* and *excellent* are retrieved and activated in the processing of SI-triggering sentences. The results of our experiments indeed suggest that alternatives play a role in the processing of SI, though a number of empirical puzzles remain.

Keywords. pragmatics; scalar implicature; alternatives; semantic priming

1. Introduction. Scalar implicature (SI), exemplified in (1), is a classic example of utterances receiving an enriched meaning that goes beyond their literal meaning.

(1) Mary ate some of the deep dish.
Literal meaning: Mary ate at least some of the deep dish.
SI-enriched meaning: Mary ate some, but not all, of the deep dish.

It is commonly assumed that the inferential process that gives rise to SI involves hearers reasoning about informationally stronger unsaid alternatives. For example, upon encountering the utterance in (1), hearers consider the stronger statement *Mary ate all of the deep dish*, which the speaker could have said, but did not. Hearers can infer as an SI the negation of this stronger alternative, enriching the literal meaning of (1) with *Mary didn't eat all of the deep dish*. This process can be viewed as an interaction of the Quality and Quantity maxims (Grice 1967).

It is an open question what psycholinguistic mechanisms underlie the inferential process that gives rise to SI. To address this, this paper uses semantic priming with lexical decision to test whether unsaid alternatives are retrieved and activated in the processing of SI-triggering sentences. The general logic of our experiments is to probe whether alternatives like *all* are recognized with facilitated reaction times in a lexical decision task when they follow a relevant SI-triggering sentence like (1). Our findings suggest that comprehenders indeed activate the alternatives that theories of SI take them to reason about; in other words, lexical scales are psychologically real.

*For helpful discussion and feedback, we would like to thank Ira Noveck and Nicole Gotzner, as well as audiences at ELM 2 and at the Laboratoire de Linguistique Formelle LingLunch. This material is based upon work supported by the National Science Foundation under Grant No. #BCS-2041312. All mistakes and shortcomings are our own. Stimuli, data, and the scripts used for data visualization and analysis can be found in the following OSF repository: https://osf.io/wga25/?view_only=cefc447bc6e649aeb4815de958d71597 Authors: Eszter Ronai, The University of Chicago (ronai@uchicago.edu) & Ming Xiang, The University of Chicago (mxiang@uchicago.edu).

©2023 Eszter Ronai and Ming Xiang
Published by the LSA with permission of the author(s) under a CC BY license.

Though much research has concentrated on the <*some*, *all*> scale and the corresponding *some but not all* SI, other lexical items also form scales. As (2) demonstrates, an utterance containing *good* can invite reasoning about the stronger alternative *excellent*, and lead to a *good but not excellent* SI. Recently, attention has turned to investigating a wider range of scales, with findings uncovering variation in the likelihood of SI: e.g., the SI in (2) is much less likely to arise than the one in (1) (i.a. van Tiel et al. 2016). In our investigation of alternative activation, we capitalize on this phenomenon of *scalar diversity*, and our priming experiments will test 60 different scales.

(2) The movie is good.
Literal meaning: The movie is at least good.
SI-enriched meaning: The movie is good, but not excellent.

This paper is structured as follows. Section 2 reviews previous work on the processing of alternatives. We then present four semantic priming experiments: Experiment 1 is a replication unrelated to SI (Section 3); Experiment 2 tests scalar alternatives without sentential context (Section 4); Experiment 3 tests alternatives in the context of SI-triggering sentences (Section 5); finally, Experiment 4 tests focus alternatives (Section 6). In Section 7, we discuss what relevance priming results can have for different theories of SI, as well as some remaining empirical puzzles.

2. Alternatives in language processing. Alternatives are pervasive in (the modeling of) semantic-pragmatic meaning. Correspondingly, they have generated a lot of interest in psycholinguistics, with various experimental paradigms being used to probe what kind of mental representations they have. In this brief section, we will concentrate on focus and scalar alternatives. For comprehensive overviews of alternative processing, including also alternatives involved in negation and counterfactuals, see Repp & Spalek (2021) and Gotzner & Romoli (2022) (and references therein).

Sentential focus marks new or emphasized information in a sentence. This information is often provided in (implicit) contrast to possible other alternatives (Rooth 1992, 1985). In English, focus can be marked for instance by placing a prominent accent on a word:

(3) Mary ate DEEP DISH.

(3) conveys not only that Mary ate deep dish, but also that Mary did not eat anything else from among a set of contextually determined of alternatives, e.g., {lasagne, salad, ...}. In successful comprehension, hearers infer this set of contrastive alternatives as intended by the speaker.

A growing number of studies have found that focus alternatives such as *lasagne* above are activated in the processing of sentences like (3). Our experiments testing scalar alternatives are modeled after studies that used semantic priming to test the processing of sentential focus and the activation of focus alternatives. In particular, Husband & Ferreira (2015) (following Braun & Tagliapietra 2009; see also Gotzner et al. 2016, Yan & Calhoun 2019) used lexical decision with cross-modal priming. In this study, participants listened to sentences such as (4), and had to make a decision about whether a visually presented target word was a word of English.

(4) The murderer killed the NURSE last Tuesday night.

The prime in each sentence was the focused element (*nurse* in (4)), while the targets in the lexical decision task were contrastive semantic associates (focus alternatives, e.g. *doctor*), non-contrastive semantic associates (e.g. *clinic*) and unrelated words. The study found early activation (i.e. facili-

tated lexical decision reaction times) of both contrastive and non-contrastive semantic associates in sentences where *nurse* was focused. Importantly, however, later activation (after a longer stimulus onset asynchrony) was only found for contrastive alternatives. This suggests that comprehenders establish the proper set of focus alternatives during comprehension, which then allows them to draw the relevant inferences intended by the speaker, i.e., that the murderer did not kill the doctor.

Though previous priming studies are the most relevant for our paper, evidence of the activation of focus alternatives comes from a much larger body of work. Existing studies have tested many different ways of marking sentential focus, including not just intonation, but focus particles (e.g., *only*, *also*), cleft sentences, or font emphasis. They have also successfully used a variety of experimental paradigms, such as probe recognition, delayed recall, change detection, or visual world eye-tracking. The readers are referred to i.a., Sanford et al. (2009), Kim et al. (2015), Fraundorf et al. (2010), Fraundorf et al. (2013), Spalek et al. (2014), and Gotzner & Spalek (2017).

Priming has also been used to investigate scalar alternatives. De Carvalho et al. (2016) used lexical decision with subliminal priming to see if one member of a scale (e.g., *some*) activates the other (*all*). Participants were visually presented with a prime word for 34ms and then had to decide whether the following visually presented target was a word of English. The authors' goal was to adjudicate between different theories of SI. They made the assumption that under a Neo-Gricean account of SI, which relies on lexically given Horn-scales, the stronger alternative *all* is always needed in the processing of the weaker term *some*, but not vice versa. This makes the prediction that *some* would prime *all* more than *all* primes *some*. A Post-Gricean account such as Relevance Theory, on the other hand, does not assign special significance to lexical scales. The authors therefore predicted that under Post-Gricean accounts, any priming effect should merely be due to semantic relatedness and not show asymmetry, i.e., *some* and *all* would prime each other equally. The findings are in line with the Neo-Gricean account. An important difference between this study and ours (as well as the literature on focus alternatives), is that de Carvalho et al. tested whether scalar terms prime each other in the absence of any sentential context. In contrast, what we are primarily interested in is whether scalar alternatives are primed in SI-triggering sentences.

Another relevant priming study is by Schwarz et al. (2016), whose research question addressed not whether scalar alternatives are activated in the processing of SI. Rather, they tested the hypothesis that by presenting an alternative as the prime, its salience is increased, which might lead to more likely and faster SI calculation. Ultimately, the findings did not support this hypothesis. Lastly, there is also a growing body of work that uses not lexical, but structural priming to investigate SI, see i.a., Rees & Bott (2018), Bott & Frisson (2022), and Bott & Chemla (2016).

3. Experiment 1: replication of Thomas, et al. (2012). Given that priming experiments are typically conducted in person in a lab setting, we first carried out a replication study of the basic semantic priming effect in Thomas et al. (2012), in order to validate our web-based methodology.

3.1. PARTICIPANTS AND TASK. 50 native speakers of American English participated in an online experiment on the PCIbex platform (Zehr & Schwarz 2018). Participants were recruited on Prolific and compensated $2. Native speaker status was established via a language background questionnaire; payment was not conditioned on the participant's response. Participants were removed if their lexical decision accuracy was below 90%. Data from 39 participants is reported below.

Experiment 1 was a semantic priming with lexical decision experiment. Participants had to

decide whether a word they saw was a word of English or not; this word was the target. They had to press the F key for "non-word" and the J key for "word". The primary dependent variable of interest is their reaction time in making this lexical decision. Participants were instructed to make a decision as fast as possible, while remaining accurate. Before making a lexical decision on the target, participants saw the prime word. There were two within-participants conditions: in the "related" condition, the target word (e.g., *boy*) was preceded by a prime word that was semantically related to it (*girl*). In the "unrelated" condition, the same target (*boy*) was preceded by a prime that was not semantically related (*boulevard*). Primes appeared in uppercase and targets in lowercase.

Participants first saw a fixation cross displayed for 350ms. It was then followed by 400ms of a blank screen. After that, the prime word appeared for 150ms. The presentation of the prime word was followed by another 650ms blank screen —this is the stimulus onset asynchrony (SOA), i.e., the time between the offset of the prime and the onset of the target. Finally, participants saw the target word, which they had to make the lexical decision on. If a participant did not make a lexical decision within 3000ms of the onset of the target, the experiment moved on to the next trial.

The related condition in Experiment 1 used 60 prime-target pairs from the "symmetrical associates" in Thomas et al. (2012; p. 640, Table A1). These pairs are symmetrical because the prime has a meaning that evokes the target, and vice versa, e.g., *girl-boy*, *circle-square*, *salt-pepper*. The unrelated prime words were randomly selected from the "forward associates" in Thomas et al. (2012; p. 640, Table A1). The experiment also included 60 fillers items with non-word targets. Of the filler targets, 30 were 4-10/11 letter pseudohomophones that we generated from the ARC Nonword Database (Rastle et al. 2002) —e.g., *spraized, knewed* —, and 30 were non-words from Lupker & Pexman (2010; p. 1282, Standards-1) —e.g., *cleam, dronk*. The experiment started with 10 practice items: 5 words and 5 non-words. For the first 4 practice items only, participants saw reminder labels that the F key corresponded to "non-word" and J to "word".

3.2. HYPOTHESIS AND PREDICTIONS. We predict to replicate Thomas et al.'s result: shorter lexical decision reaction times (RT) in the related, as compared to the unrelated condition. In the related condition, the target has been preceded by a semantically similar word, which should activate its meaning and facilitate its recognition. In the unrelated condition, the prime would not activate the target, which is then recognized at a "baseline" speed, related to its frequency, length, etc. (See also i.a., Swinney 1979, Swinney et al. 1979 for classic findings of semantic priming.)

3.3. RESULTS AND DISCUSSION. Data points with incorrectly answered lexical decision responses (i.e., a "non-word" response) were excluded, removing 2.09% of the data. Figure 1 shows mean RT (and standard error) by condition. For the statistical analysis, a linear mixed effects regression model was fit (lme4 package in R; Bates et al. 2015), predicting RT on the target word by Condition ("related" vs. "unrelated"). The fixed effects predictor Condition was sum-coded (related: -0.5 and unrelated: 0.5). Random intercepts and slopes were included for participants and items. RTs in the related condition were found to be significantly shorter than in the unrelated condition (Estimate=25.51, SE=8.65, t=2.95, $p<0.01$). That is, participants recognized words faster when they have been primed by a semantically related word.

This successfully replicates Thomas et al.'s in-lab results, validating the web-based setup —though we must note that the magnitude of the priming effect (i.e., the difference in RT between the related and unrelated conditions) was smaller in our experiment.

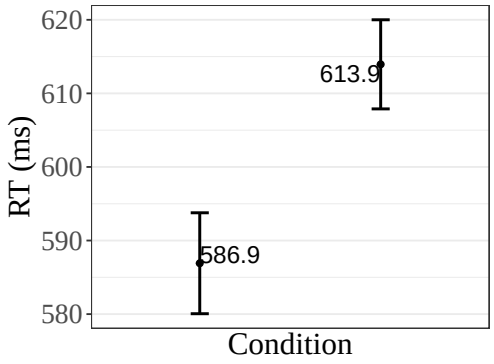

Figure 1: Results of Experiment 1: replication of Thomas et al. (2012)

4. Experiment 2: lexical priming. Experiment 2 was also a lexical semantic priming experiment, but this time the primes were weaker scalar terms from a scale (e.g., *some, good*), while the targets were stronger alternatives (e.g., *all, excellent*). Importantly, scalar term primes were not placed in a sentential context, where SI could have been calculated. This means that if we see semantic priming in Experiment 2, that will reflect semantic similarity, not SI calculation. Experiment 2 therefore provides a baseline for later experiments that test inference-triggering sentences.

4.1. PARTICIPANTS AND TASK. 49 native speakers participated for $2 compensation. Recruitment and screening was identical to Experiment 1, including the exclusion criterion. Data from 44 participants is reported below.

Capitalizing on scalar diversity, Experiment 2 used 60 different lexical scales as critical items. Each item consisted of a pair of scalar terms where the stronger term asymmetrically entails the weaker one —see Ronai & Xiang (to appear) for how this scale set was constructed. Prime words in the "related" condition were weaker scalar terms like *good*, while targets were stronger alternatives like *excellent*. In the "unrelated" condition, the primes were instead unrelated words such as *foreign*. Unrelated primes were generated to satisfy two criteria. First, they had to fit into the sentence frames used in Experiments 3-4: given the sentence *The movie is good*, *foreign* was chosen, since *The movie is foreign* is also an acceptable sentence. Second, unrelated primes had to have sufficiently low semantic similarity with the target (average cos(θ)=0.138). This was operationalized using vector semantics, specifically the GLoVe model and spaCy word embeddings.

Other than the critical test items, Experiment 2 was identical to Experiment 1 in its task, procedure (including timing parameters such as SOA), filler and practice items.

4.2. HYPOTHESIS AND PREDICTIONS. If pairs of scalar terms (e.g., *good-excellent*) are semantically similar enough to lead to priming, then the results of Experiment 2 should pattern similarly to Experiment 1 —we should see shorter RTs in the related condition than in the unrelated condition.

4.3. RESULTS AND DISCUSSION. Data points with incorrectly answered lexical decision responses ("non-word") were excluded, removing 2.35% of the data. Figure 2 shows mean RT (and standard error) by condition. Statistical analysis was identical to Experiment 1, except for the random effects structure, which included random intercepts for participants and items and random slopes for participants. The statistical analysis revealed no significant difference between RTs in

the related and unrelated conditions (Estimate=11.46, SE=9.94, t=1.15, p=0.26).

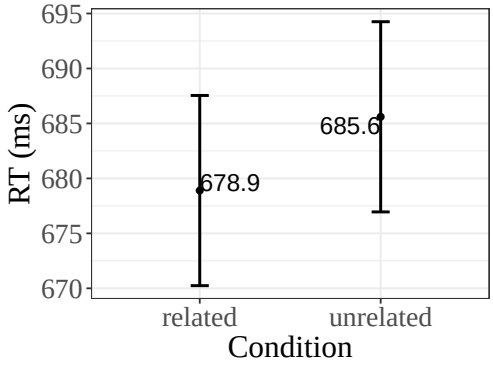

Figure 2: Results of Experiment 2: lexical priming experiment testing scalar alternatives

That is, targets in the related condition were not recognized significantly faster than in the unrelated condition. This suggests that pairs of scalar terms do not lead to semantic priming when the words are presented in isolation, in the absence of any sentential context. Therefore, we will be able to conclude that any priming effect we find in sentential experiments (Experiments 3-4) is due to inference processing and alternative retrieval, not just mere meaning similarity.

5. Experiment 3: sentential priming. Having seen in Experiment 2 that weaker scalar terms do not prime stronger alternatives in isolation, Experiment 3 turns to priming effects in sentential contexts. Here, we test whether stronger scalar alternatives are retrieved and activated in the processing of sentences that lead to SI calculation, and are taken to involve reasoning about alternatives.

5.1. PARTICIPANTS AND TASK. 50 native speakers participated for $3.20/3.50 compensation. Recruitment and screening was identical to Experiment 1. Data from 46 participants is reported.

Experiment 3 was also a lexical decision task with two within-participants conditions (related vs. unrelated). Target words were the same scalar terms as in Experiment 2. Importantly, however, primes were now full sentences: the prime words from Experiment 2 were embedded in a sentential context. That is, while for the <*good*, *excellent*> scale Experiment 2 used the word *good* as a prime, in Experiment 3 *good* appeared in a sentence: *The movie is good*. Similarly, in the unrelated condition, the unrelated words were embedded in a sentential context, e.g., *The movie is foreign*.

Each trial started with a 350ms fixation cross. This was followed by 400ms of a blank screen. After that, prime sentences were presented word-by-word, with each word being displayed for 350ms. There was a 650ms SOA between the offset of the final word in the sentence (*good/foreign*) and the onset of the target (*excellent*). As before, if a lexical decision was not made within 3000ms of the onset of the target, the experiment moved on to the next trial. Filler and practice targets used the materials of Experiments 1-2, but the primes were sentences, not single words.

5.2. HYPOTHESIS AND PREDICTIONS. If stronger scalar alternatives like *excellent* are reasoned about and retrieved during SI calculation, then we should see shorter RTs in the related condition than in the unrelated condition. That is, *excellent* should be recognized faster following an SI-triggering sentence where it serves as a stronger alternative. On the contrary, if lexical alternatives do not play a role in SI processing, then there should be no RT difference between the conditions.

5.3. RESULTS AND DISCUSSION. Excluding incorrectly answered lexical decision responses ("non-word") removed 1.7% of the data. Figure 3 shows mean RT (and standard error) by condition. Statistical analysis was identical to Experiment 2. RTs in the related condition were found to be significantly shorter than in the unrelated condition (Estimate=21.62, SE=8.65, t=2.5, p<0.05); targets were recognized faster following an SI-triggering sentence.

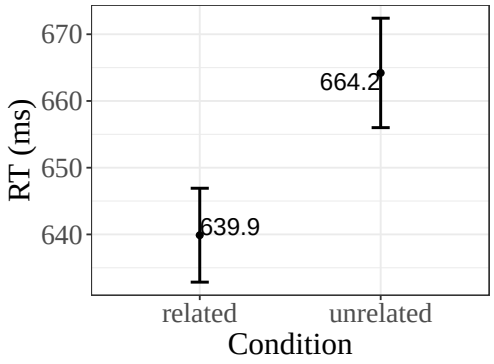

Figure 3: Results of Experiment 3: sentential priming experiment testing scalar alternatives

Experiment 3's findings therefore show that a stronger scalar alternative like *excellent* is recognized faster as a word of English when it has been preceded by a sentence like *The movie is good*, which can trigger the *not excellent* SI. This, in turn, suggests that in the processing of such an SI-triggering sentence, comprehenders retrieved and activated the relevant stronger scalar alternative. In the unrelated condition, on the other hand, such alternative targets would not have been activated in the processing of the prime sentence, and were therefore recognized with a baseline RT. Let us recall that these findings cannot receive an explanation simply in terms of semantic similarity. The prime sentences were identical across the related and unrelated conditions up until the critical word (*The movie is X*). And as for the critical word (the weaker scalar term *good* vs. the unrelated word *foreign*), Experiment 2 demonstrated that their difference in meaning, and the similarity between *good* and *excellent* does not, in itself, lead to semantic priming.

There is, however, one important caveat: we cannot be certain that what the priming effect is evidence for is the retrieval of specific lexical items (e.g., *excellent*). It is also possible that the observed facilitation in RTs is due to a more general activation of semantic features associated with the stronger alternative state. For instance, given the sentence *The movie is good*, participants might have considered the stronger alternative state where the movie is more than good, but without necessarily reasoning about the specific alternative *excellent* —this might still result in the observed effect. We also cannot be certain that participants in Experiment 3 actually calculated the SIs (e.g., *not excellent*), since the experiment did not include a task to probe SI calculation.

6. Experiment 4: sentential priming with *only*. Numerous studies have shown that focus alternatives are activated in sentence processing (Section 2). As another baseline to Experiment 3, we therefore conducted an experiment where prime sentences also included the focus particle *only*.

6.1. PARTICIPANTS AND TASK. 50 native speakers participated for $3.20 compensation. Recruitment and screening was identical to Experiment 1. Data from 43 participants is reported below.

Experiment 4 was identical to Experiment 3 in its task and procedure (including timing param-

eters), with one difference: critical items were modified such that prime sentences in the related condition also included the focus particle *only*. That is, before participants made a lexical decision on a stronger alternative target such as *excellent*, they saw the prime sentence *The movie is only good* (presented word-by-word). The unrelated conditions, as well as filler and practice items were identical to Experiment 3, i.e., they were not modified to include the word *only*.

6.2. HYPOTHESIS AND PREDICTIONS. The exclusion of focus alternatives is encoded semantically (Rooth 1985, 1992), while alternatives in SI are excluded pragmatically, in a cancellable way. Given that Experiment 3 already revealed evidence for alternative activation in SI, we can predict to see similar effects in Experiment 4. Moreover, this is also what we expect based on previous work that has tested focus alternatives in a variety of experimental paradigms. This leads to the strong prediction that RTs should be shorter in the related condition than the unrelated condition.

6.3. RESULTS AND DISCUSSION. Excluding incorrectly answered lexical decision responses ("non-word") removed 1.98% of the data. Statistical analysis was identical to Experiment 2. Figure 4 shows mean RT (and standard error) by condition. RTs in the related condition were significantly shorter than in the unrelated condition (Estimate=24.47, SE=8.01, t=3.06, p<0.01).

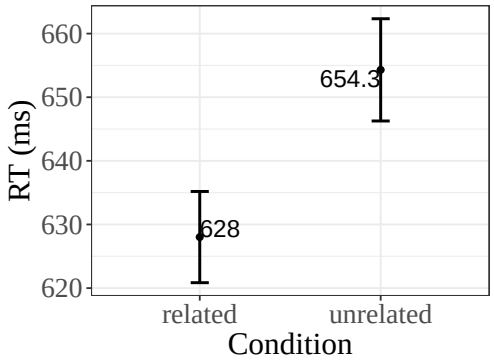

Figure 4: Results of Experiment 4: sentential priming experiment testing focus alternatives

Like Experiment 3, Experiment 4 revealed facilitation for stronger alternative targets in the related condition. That is, the prime sentence *The movie is only good* led to a faster recognition of the word *excellent*. Semantic theory holds that sentences including focus (signalled e.g., by *only*) encode the exclusion of alternatives —our findings provide evidence that excluded alternatives have a processing correlate. This is in line with existing work reviewed in Section 2.

7. Summary of findings. The experiments in this paper (especially Experiment 3) show evidence that stronger scalar alternatives (*all*, *excellent*) are retrieved and activated in the real-time processing of SI. This informs our understanding of the mental representations behind pragmatic reasoning. Classic Gricean accounts of SI hold that comprehenders reason about, and derive the negation of, relevant informationally stronger alternatives that the speaker could have said, but did not say. Our findings suggest that this reasoning process also has processing correlates: relevant alternatives are activated when hearers process SI-triggering sentences. In what follows, we discuss to what extent our findings might further inform theory, as well as some remaining puzzles.

7.1. RELEVANCE FOR THEORY?. Previous work on the activation of scalar alternatives has used findings to adjudicate between Neo- and Post-Gricean accounts (de Carvalho et al. 2016). Here, we briefly review these accounts and discuss what relevance priming evidence might have for them.

Neo-Gricean accounts typically assume that hearers infer the negation of informationally stronger alternatives that the speaker could have said —e.g., because <*some, all*> form a lexical scale, and *all* is stronger than *some*, hearers derive *not all* upon encountering *some*. These alternatives are determined via the lexicon (i.a., Horn 1972) or the grammar (i.a., Katzir 2007). On the other hand, Post-Gricean accounts such as Relevance Theory (i.a., Sperber & Wilson 1995) do not attach special importance to lexical scales. Instead, the final interpretation of an utterance is obtained through a context-based enrichment process or ad hoc concept construction (in the case of SI, strengthening). One seemingly straightforward way to interpret the findings of our experiments, then, is that they support Neo-Gricean accounts of SI, since those take hearers to reason about particular lexical alternatives. We could also say that our results are not predicted by theoretical accounts of SI that dispense with lexical scales, such as Relevance Theory. However, we believe that at least two issues arise with this interpretation of the results, having to do with what predictions different theories of SI may make for priming data.

First, it is not clear whether Neo-Gricean accounts would predict that stronger alternatives are activated when the weaker scalar term is presented in isolation. One the one hand, we could assume that lexical scales should only be relevant in language processing when alternatives are actually reasoned about, in the context of an SI-triggering utterance. If so, then the findings of this paper do indeed support Neo-Gricean accounts, since we found priming in a sentential context (Experiment 3), but not in isolation (Experiment 2). On the other hand, if lexical scales are hardwired into the lexicon, then we might predict that pairs of scalar terms prime each other even in the absence of an SI-triggering sentence. As mentioned, de Carvalho et al. (2016) made a prediction along these lines: that the weaker scalar term primes the stronger alternative asymmetrically. Following this reasoning, our findings in fact do not fully support Neo-Gricean accounts, since no priming was found when the scalar terms occurred in isolation (Experiment 2). The finding that alternatives are only activated in (a sentential) context (Experiment 3) could even be argued to support Relevance Theory, where SI calculation occurs only when there is sufficient support from context.

Second, the activation of alternatives may signal either that alternatives were retrieved for the SI calculation process to occur, or we might see activation as a by-product of the SI calculation process. Broadly speaking, Neo-Gricean accounts would assume that for SI to arise, particular lexical items from lexical scales are reasoned about —this would predict the priming effect we found in Experiment 3. But it is also possible that the priming effect we see is epiphenomenal. On Post-Gricean theories, hearers still calculate SI, even though lexical scales do not play a special role. And once hearers have reached the SI-enriched meaning (≈*The movie is no more than good.*), this could then lead to the observed priming effect, even if the stronger alternative *excellent* was not retrieved in the first place. These two ways of interpreting the priming findings are related to the issue discussed in Section 5.3: whether facilitated RTs constitute evidence that a specific lexical item *excellent* was retrieved, or whether they simply suggest that some semantic features related to the alternative state "the movie is more than good" were activated.

Given the above, we would argue that as things stand, no firm conclusions can be reached

about the validity of Neo-vs. Post-Gricean accounts based on priming evidence.

7.2. REMAINING EMPIRICAL PUZZLES. Some open questions remain. First, we saw that Experiments 3 and 4 pattern alike: RTs were facilitated in the related condition. An additional statistical analysis on the combined Experiment 3-4 data set also found no effect of Experiment (Estimate=9.51, SE=22.53, t=0.42, p=0.67). This suggests that alternatives like *excellent* are similarly activated no matter whether the sentence that is processed is *The movie is good* or *The movie is only good*. This is despite the fact that SI excludes alternatives pragmatically, but in the case of focus, alternative exclusion is semantic. Indeed, *The movie is only good* is significantly more likely to lead to the *not excellent* inference than *The movie is good* (Ronai & Xiang 2022). The lack of a difference between the current Experiment 3 and 4 suggests that the activation of alternatives, as measured via priming, does not track the rate of inference from the corresponding sentences: more robust inference calculation does not correspond to stronger priming.

Second, Experiment 2 served to rule out the possibility that a priming effect in Experiment 3 would reflect mere meaning similarity, rather than the processing correlate of reasoning about alternatives. For this reason, it is a welcome result that Experiment 2 revealed no effect. But it is itself a puzzle why we found semantic priming in Experiment 1 but not in 2. Differences in (vector) semantic similarity cannot provide an explanation: Table 1 shows that average prime-target similarity (based on GLoVe/spaCy) in the related vs. unrelated condition is quite similar in the two experiments. It is perhaps possible that the nature of the relationship between prime and target is different in Thomas et al. (2012) (and classic semantic priming studies) than in the case of scalar items: there is an intuitive sense in which *girl-boy* and *salt-pepper* are closer semantically than *good-excellent* and *some-all*. Future research should probe the exact source of semantic priming.

Cosine similarity	**Related condition**	**Unrelated condition**
Experiment 1 (replication of Thomas et al.)	0.605	0.126
Experiment 2 (scalar items in isolation)	0.707	0.138

Table 1: Average semantic similarity between prime-target pairs in Experiments 1 and 2.

Lastly, as mentioned in the Introduction, likelihood of SI varies across lexical scales. This, however, does not correspond to a systematic difference in priming. To test this, we looked at the correlation between the Experiment 3 priming effect across items and the corresponding SI calculation rates (from Ronai & Xiang to appear, Experiment 1) —but found no significant effect (Pearson's correlation test: r=0.004, p =0.98). One possible reason is that the priming effect is measured by comparison with the unrelated condition (RT on *excellent* given *The movie is good* vs. *The movie is foreign*.) The "unrelated" words (here, *foreign*) might introduce variation across items, which might obscure a potential by-item effect. Future work should use a more uniform unrelated condition, e.g. identity (*The movie is excellent*) or antonym priming (*The movie is bad*).

8. Conclusion. In this paper, we provide evidence from semantic priming that scalar alternatives (*excellent*) are activated in the processing of the relevant SI-triggering sentences (*The movie is good*). This suggests that scalar alternatives pattern similarly to focus alternatives. At the same time, a number of empirical puzzles remain for future work, and we have argued for caution in using priming evidence to draw strong conclusions about pragmatic theory.

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. 10.18637/jss.v067.i01.

Bott, Lewis & Emmanuel Chemla. 2016. Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language* 91. 117–140. 10.1016/j.jml.2016.04.004. New Approaches to Structural Priming.

Bott, Lewis & Steven Frisson. 2022. Salient alternatives facilitate implicatures. *PLOS ONE* 17(3). 1–10. 10.1371/journal.pone.0265781.

Braun, Bettina & Lara Tagliapietra. 2009. The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes* 25(7-9). 1024–1043. 10.1080/01690960903036836.

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. 10.3389/fpsyg.2016.01500.

Fraundorf, Scott H., Aaron S. Benjamin & Duane G. Watson. 2013. What happened (and what did not): Discourse constraints on encoding of plausible alternatives. *Journal of Memory and Language* 69(3). 196–227. 10.1016/j.jml.2013.06.003.

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. 10.1016/j.jml.2010.06.004.

Gotzner, Nicole & Jacopo Romoli. 2022. Meaning and alternatives. *Annual Review of Linguistics* 8(1). 213–234. 10.1146/annurev-linguistics-031220-012013.

Gotzner, Nicole & Katharina Spalek. 2017. Role of contrastive and noncontrastive associates in the interpretation of focus particles. *Discourse Processes* 54(8). 638–654. 10.1080/0163853X.2016.1148981.

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. 10.1017/langcog.2015.25.

Grice, Herbert Paul. 1967. Logic and Conversation. In Paul Grice (ed.), *Studies in the Way of Words*, 41–58. Harvard University Press.

Horn, Laurence R. 1972. *On the semantic properties of logical operators in English*: UCLA dissertation.

Husband, E. Matthew & Fernanda Ferreira. 2015. The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience* 31(2). 217–235. 10.1080/23273798.2015.1083113.

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

Kim, Christina S., Christine Gunlogson, Michael K. Tanenhaus & Jeffrey T. Runner. 2015. Context-driven expectations about focus alternatives. *Cognition* 139. 28–49. 10.1016/j.cognition.2015.02.009.

Lupker, Stephen J. & Penny M. Pexman. 2010. Making things difficult in lexical decision: the impact of pseudohomophones and transposed-letter nonwords on frequency and semantic prim-

ing effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36(5). 1267–1289. 10.1037/a0020125.

Rastle, Kathleen, Jonathan Harrington & Max Coltheart. 2002. 358,534 nonwords: The arc nonword database. *The Quarterly Journal of Experimental Psychology Section A* 55(4). 1339–1362. 10.1080/02724980244000099. PMID: 12420998.

Rees, Alice & Lewis Bott. 2018. The role of alternative salience in the derivation of scalar implicatures. *Cognition* 176. 1–14. 10.1016/j.cognition.2018.02.024.

Repp, Sophie & Katharina Spalek. 2021. The role of alternatives in language. *Frontiers in Communication* 6. 10.3389/fcomm.2021.682009.

Ronai, Eszter & Ming Xiang. 2022. Quantifying semantic and pragmatic effects on scalar diversity. In *Proceedings of the Linguistic Society of America 7(1)*, .

Ronai, Eszter & Ming Xiang. to appear. Three factors in explaining scalar diversity. In Daniel Gutzmann & Sophie Repp (eds.), *Proceedings of Sinn und Bedeutung 25*, .

Rooth, Mats. 1985. *Association with focus*: University of Massachusetts, Amherst dissertation.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. 10.1007/bf02342617.

Sanford, Alison J. S., Jessica Price & Anthony J. Sanford. 2009. Enhancement and suppression effects resulting from information structuring in sentences. *Memory & Cognition* 37(6). 880–888. 10.3758/mc.37.6.880.

Schwarz, Florian, Jérémy Zehr, Daniel Grodner & Hezekiah Akiva Bacovcin. 2016. Subliminal priming of alternatives does not increase implicature responses. Poster presented at the Logic and Language in Conversation Workshop, University of Utrecht.

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. 10.1016/j.jml.2013.09.001.

Sperber, Dan & Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Wiley-Blackwell 2nd edn.

Swinney, David A. 1979. Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18(6). 645–659. 10.1016/S0022-5371(79)90355-4.

Swinney, David A, William Onifer, Penny Prather & Max Hirshkowitz. 1979. Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition* 7(3). 159–165. 10.3758/bf03197534.

Thomas, Matthew A., James H. Neely & Patrick O'Connor. 2012. When word identification gets tough, retrospective semantic processing comes to the rescue. *Journal of Memory and Language* 66(4). 623–643. 10.1016/j.jml.2012.02.002.

van Tiel, Bob, Emiel Van Miltenburg, Natalia Zevakhina & Bart Geurts. 2016. Scalar diversity. *Journal of Semantics* 33(1). 137–175. 10.1093/jos/ffu017.

Yan, Mengzhu & Sasha Calhoun. 2019. Priming effects of focus in mandarin chinese. *Frontiers in Psychology* 10. 10.3389/fpsyg.2019.01985.

Zehr, Jeremy & Florian Schwarz. 2018. PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832.