C O P *i* L CAMBRIDGE OCCASIONAL PAPERS *in* LINGUISTICS
Volume **12**, Article 2: pp. 31–46, 2020 | ISSN: 2050-5949

An Analysis of Scalar Items under Memory Load Using an Inference Task*

Blanche Gonzales de Linares
University of Cambridge

Abstract This paper explores the processing of classic scalar terms such as *some* and compares them with both numerals and scalar adjectives. Indeed, previous empirical research on *some* has suggested that scalar implicatures are cognitively costly and that with all scalars when cognitive resources are being used for another task participants will revert to their logical *at least* reading. However, theoretical research has suggested that numbers could in fact have an exact logical reading, a claim which is supported by empirical findings. Moreover, recent research also suggests that for some scalar adjectives the stronger scale mate does not entail the weaker one. This could mean that these scalar adjectives in fact have an exact primary meaning that is pragmatically enriched to an *at least* reading. Using a memory load task combined with an inference task, the present experiment provides additional support for the exact theory of numbers, as well as for the aforementioned theory that some scalar adjectives could be processed differently than classically studied scalars.

1 Introduction

In purely logical terms, the meaning of *some* is compatible with that of *all*. Indeed, *all* entails *some*, and a speaker saying (1a) when (1b) is true would logically be correct: if all of the students went, it therefore means that some went.

(1) a. Some of the students went to the concert.
b. All of the students went to the concert.
c. Some but not all of the students went to the concert.

However, in conversation, (1a) is often understood to mean (1c), which is incompatible with (1b). This is what is called a scalar implicature: the speaker is following Grice's maxim of quantity to "make your contribution as informative as is required" (Grice 1975). Indeed, if the speaker believed *all* to be true, they would not use the weaker scale-mate *some*, as it is less informative, which violates the maxim of quantity. The semantically encoded meaning of *some* is therefore *at least some*,

* This paper is an abridged version of my long essay for the BSc in Experimental Linguistics at University College London. I want to thank my supervisor, Dr Richard Breheny, for his advice and feedback, as well as Dr Chao Sun for her technical help.

and is then pragmatically enriched by the scalar implicature to mean *some but not all*.

This phenomenon is not limited to the ⟨some, all⟩ scale, but can be applied to various types of scalar terms, such as adjectives (⟨intelligent, brilliant⟩), adverbs (⟨sometimes, alway⟩), or nouns (⟨mammal, dog⟩) (Van Tiel, Van Miltenburg, Zevakhina & Geurts 2016).

2 Cognitive Processing of Scalar Implicatures

An important question in the study of scalar implicatures is whether they are cognitively costly. One view holds that implicatures are automatic (Levinson 2000). In such a view, the *some but not all* reading is processed by default. The meaning of *some* that is compatible with *all* can then be derived through the cancellation of the implicature, which requires additional cognitive processing. The opposite view, the context driven view, holds that implicatures are only derived when there is contextual support: the hearer would therefore start by processing the logical *at least some* meaning, and would then derive the implicature if the context supports it, which leads to an added processing cost (Carston 2011, Sperber & Wilson 1986).

Experimental research testing this phenomenon has shown that the processes underlying scalar implicature seems to be cognitively demanding, which supports the context-driven view (Bott & Noveck 2004, Breheny, Katsos & Williams 2006, De Neys & Schaeken 2007). The experiment most relevant to the current study is De Neys & Schaeken's (2007) experiment, which shows that under memory load participants tend to revert to the logical reading of *some*. This suggests that implicature is cognitively costly as the memory task is using the same resources as the implicature derivation, which leads to the lower rates of implicature.

This conclusion has been somewhat questioned by a recent study, which suggests that this effect could be due to the fact that ⟨some, all⟩ is positively scalar, which leads to the introduction of a lower bound and requires the processing of negative information, which has been shown to be cognitively costly for several reasons, such as the possible presupposition that the positive counterpart is true (Moxey 2006). This would predict that increasing the memory load only leads to more logical readings with positively scalar expressions such as ⟨some, all⟩ or ⟨might, will⟩, but not with negatively scalar ones such as ⟨low, empty⟩ or ⟨scarce, absent⟩. The results confirm this prediction, as only positively scalar expressions were significantly impacted by memory load (Van Tiel, Pankratz & Sun 2019). When discussing the cognitive cost of scalar inference, therefore, it is important to bear in mind that this effect could be due to the negative information processed during the inference rather than the inference itself as previously claimed by De Neys & Schaeken (2007). Crucially, however, for positive scales the prediction made by both theories is that increasing memory load will decrease the rates of inference. As in this experiment I will only be testing positive scales, I will therefore put this issue to the side and simply assume that there is an undetermined processing cost involved. Whatever the reason for the processing cost, in both cases the effect of memory load should cause the speaker to revert to the primary logical meaning of the scalar expression, such as *at least some* for the ⟨some, all⟩ scale, which is what is relevant to the present study.

3 Processing of Number Scales

A possible claim that could be made at this point is that, with all positive scales, deriving an exact meaning from an *at least* one will be cognitively costly. Similar patterns of results should therefore be expected with all these positive scalars tested under memory load (Van Tiel et al. 2019). However, research on numbers has shown that they do not follow that pattern. It is important to note that numbers do not have an upper bound, as they are an infinite scale. However, it can be argued that the scalar inference that is derived from a number term negates the next higher scale-mate, or even all higher scale-mates. Following the Neo-Gricean method, *not five* or *not more than four* would be derived from the number *four*, which introduces negation. The predictions made above regarding the processing cost of positive scalars should therefore apply to numbers. However, this does not seem to be the case. Marty, Chemla & Spector (2013) compared ⟨some, all⟩ and numbers with a graded sentence-picture matching task paired with a memory task. For ⟨some, all⟩, the results replicated the results found by De Neys & Schaeken (2007). With numerals, however, the effect was reversed. Indeed, under low memory load, participants preferred the *at least* reading of numerals, and under high memory load they preferred the exact reading. As numbers are a positive scale, this is inconsistent with the prediction suggested above that all positive scalars are processed the same. Following the assumption that people are more logical under memory load, the results for numbers have interesting theoretical implications.

Indeed, there has been much theoretical debate about the primary reading of numbers. The traditional Neo-Gricean theory claims that the semantically encoded meaning of numbers is an *at least* reading, and that this meaning is then pragmatically enriched through a scalar implicature to create an exact one (Horn 1972). In the same way that *some* comes to mean *some but not all* through Grice's maxim of quantity, a number *n* would have *at least n* as its semantically encoded reading, and since the speaker would have used a higher scale-mate if it were more informative, the numerically quantified expression is then enriched to mean *exactly n* (Grice 1975, 1968 as cited by Horn, unpublished as cited by Horn, Horn 1972).

The assumption that the *at least* meaning is primary has since then been challenged. One theory posits a primary exact meaning of numbers that is then pragmatically enriched to lead to *at least* interpretations (Breheny 2008). Another prominent theory is based on the assumption that numbers are ambiguous between their *at least* interpretation and their exact one. It matches the exact theory of numbers, however, in the claim that the exact meaning is the primary one, although it claims that the *at least* reading is derived through semantic type-shifting rules rather than pragmatically, through the successive uses of Quantified Lowering and Existential Closure (Geurts 2006).

These theories have been supported by developmental evidence. Indeed, children have been shown to have difficulties with deriving scalar inference, and to prefer the logical reading of scalar terms such as *some* (Noveck 2001). When tested with numbers, however, they showed adult-like responses and preferred the exact reading (Papafragou & Musolino 2003). If children prefer logical semantically encoded readings to pragmatically enriched ones, the fact that they interpret numbers in their exact reading suggests that the underlying semantic meaning of numbers could be the exact rather than the *at least* one.

The developmental and memory load studies both support the view that numbers have an exact primary meaning that then becomes an *at least* meaning though semantic or pragmatic processes. The present study will aim to further explore this question by replicating the Marty et al. (2013) results with a different task.

4 Processing of Scalar Adjectives

Another interesting question concerns the less-studied scales such as scalar adjectives. Indeed, the majority of the experimental research on scalars focuses on ⟨some,all⟩, or on a very limited number of other scales (Van Tiel et al. 2016). While various scalar terms have been shown to lead to varying rates of inference (Van Tiel et al. 2016), it is commonly assumed that they are all different degrees of the same processing mechanism. ⟨some, all⟩ is therefore considered to be representative of all scales, excluding numbers due to the debate about their underlying semantics. It could be, however, that this is not the case. Sun, Tian & Breheny (2018) have tested scalar terms using a task, the 'so-task', which tests for the likelihood of a term undergoing local enrichment to exclude situations where the stronger scale-mate is true.

(2) The water is hot so not warm. (Sun et al. 2018)

Example 2 suggests that one might expect *not warm*, given *hot*. *Warm*, however, is a lower scale-mate than *hot*. This paper started with the observation that *all* entails *some*, as it is higher on the same scale. Indeed, (3b) entails (3a). If a scale has high naturalness ratings on the so-task, such as ⟨warm, hot⟩, however, it means that a version of a sentence containing the stronger scale-mate does not mean that the same sentence containing the weaker scale-mate is true.

(3) a. The water is hot.

b. The water is warm.

This means that (3a) does not entail (3b). This leads to the question of how scales that are rated high on the so-task are processed. Indeed, ⟨warm, hot⟩ has a high rating on the inference task, but it could be that this result is due to a different processing mechanism than scales such as ⟨possible, certain⟩ which also have high ratings on the inference task but have low rates on the so-task. Scales like ⟨warm, hot⟩ could in fact be processed similarly to the exact view of numbers, in which case *warm* would mean something like *exactly warm* in its primary sense, in the same way *three* means *exactly three*. Some form of pragmatic enrichment would then lead to the *at least warm* sense, similarly to numbers in the exact theory of numerals. If this is the case, similar patterns should be expected for numbers and ⟨warm, hot⟩ under memory load, in contrast to the patterns found for ⟨some, all⟩ and ⟨possible, certain⟩.

5 Design and Predictions of the Present Study

The present study uses a new method to test the various types of scalar listed above under memory load. Indeed, the task used by De Neys & Schaeken (2007) can only be used for *some*, as it tests participants using a noun phrase that is part of the category described by another noun phrase, such as *some tuna are fish*, in which *tuna* is contained within the category *fish*. The tasks used in Van Tiel et al. (2019) and Marty et al. (2013) both use visual tasks, which enables a wider variety of scales to be tested but limits the research possibilities to scales that can easily be visually represented. The present study tries to widen the field of possible scales by combining an inference task that can use any type of scalar word with a classic working memory load task. The inference task is inspired by the one used byVan Tiel et al. (2016) to test effects of scalar diversity.

An example material from Van Tiel et al. (2016) is shown in Figure 1. The inference task was used to test effects of diversity in the rate of scalar implicatures with various scalar expressions. Although that issue is not the focus of the present paper, it enabled us to choose three scales that were shown to have high rates of inference with the inference task: ⟨some, all⟩, ⟨possible, certain⟩ and ⟨warm, hot⟩. This is important, as scales with low rates of inference on the inference task would be less likely to show a decrease of inferences under memory load, and more generally would be less well suited to represent the phenomenon of scalar implicature.

John says:

She is intelligent

Would you conclude from this that, according to John, she is not brilliant?

☐ Yes ☐ No

Figure 1 Example material from Van Tiel et al. (2016).

The choice of scales was motivated by several reasons. First of all, ⟨some, all⟩ is the most commonly tested scale (Van Tiel et al. 2016), and as it is a quantifier it is the scale most similar in its use to numbers. It therefore stands to reason that it should be included in the experiment, as a way to attempt to replicate previous results and to compare efficiently with numbers.

The ⟨possible, certain⟩ scale was chosen to replicate the results found for ⟨might, will⟩ in Van Tiel et al. (2019), as *it is possible that n will m* is very close in meaning to *n might m*. Indeed, the results for that scale were clearer than the results for *some*, which makes it a good candidate to see an effect of memory load. ⟨possible, certain⟩ was used instead of ⟨might, will⟩ because it is a better item to compare with

⟨warm, hot⟩. Indeed, both ⟨possible, certain⟩ and ⟨might, will⟩ have similar results in the so-task (⟨possible, certain⟩ is actually rated lower), but ⟨possible, certain⟩ are adjectives, which removes a potential factor of word type for the comparison with ⟨warm, hot⟩. Numbers and ⟨warm, hot⟩ were chosen to explore the theoretical questions discussed above.

Several predictions can be sketched out. First of all, an effect of load on the ⟨some, all⟩ and ⟨possible, certain⟩ scales is expected, which would replicate previous results. For these scales, as the cognitively effortful process is predicted to be the one deriving *some but not all* from *at least some*, I expect that there will be an increase in difficulty on the memory load task to lead to a reduced amount of yes answers on an inference task such as the one presented in Figure 1. Getting this result would show that such an inference task can be used to find an effect of memory load, which is my first aim.

For numbers and ⟨warm, hot⟩, however, the opposite results are expected. Indeed, my theoretical prediction is that they are processed similarly to each other, and that they have a primary exact reading which is then pragmatically enriched to have an *at least* reading, in which case an increase in memory load would lead participants to revert to the logical exact meaning, which is indicated by high rates of yes responses in an inference task like the one in Figure 1.

A general prediction is also that numbers and ⟨warm, hot⟩ will be processed the same under low load, and that those scales will be processed differently than ⟨some, all⟩ and ⟨possible, certain⟩. The two latter scales should be processed the same. This would bring support to the theory that there are two different processes at hand in the processing of these two categories of scales. The rates of yes for ⟨warm, hot⟩ and numbers under memory load should be lower than the rates for ⟨some, all⟩ and ⟨possible, certain⟩ under low load, as their pragmatically enriched meaning should be the *at least* one, which is indicated by a lower rating of *yes* in the inference task.

6 Experiment

6.1 *Participants*

Participants were recruited online through the Prolific Academic platform. 80 native speakers of English, ranging in age from 18 to 65, participated in the experiment. Participants were compensated $1 for their time.

6.2 *Materials and Tasks*

6.2.1 *Inference Task*

The main task was an inference task based on the materials of Van Tiel et al. (2016), an example of which is presented in Figure 1. Participants read a statement said by a character, followed by a yes/no question about that statement, that they had been instructed to answer using two keys on their keyboard. Characters were called either John, Mary, Jane or Peter. The statement contained a scalar expression. The question asked the participants whether, according to the speaker, the statement implied that the same statement with a stronger scale mate as the scalar expression would have been false. An example of a target item is shown in Figure 2 below.

Peter says:

Some of the children are smiling.

Would you conclude from this that, according to Peter,

not all of the children are smiling?

1:Yes 2:No

Figure 2 Example of a target item.

There were two target items for each of the scales (*some/all, possible/certain, warm/hot* and numbers). For each scale there were also two true control items, where the statement in the question was undeniably compatible with the character's original statement, and two false control items, where the two statements were undeniably incompatible. A full list of items can be found in Appendix A.

6.2.2 Memory Task

The memory task was identical to the one used by Van Tiel et al. (2019). Before each item of the inference task, participants were presented with a 3x3 grid with a pattern of black cells and asked to memorize it. In the low load condition, the pattern was three black cells in a horizontal or vertical straight line. In the high load condition four cells were black, and they were scattered across the grid. After the inference task, participants were asked to replicate that grid. An example of the memory task is shown in Figure 3.

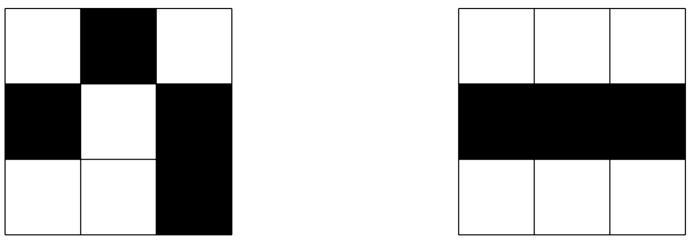

Figure 3 Examples of low-load and high-load grids.

6.2.3 Procedure

The dual-task procedure is illustrated in Figure 4. At the beginning of each trial, participants were shown the grid for 850 ms. Then, an inference task item was displayed on the screen and remained until the participant had answered the question. Participants were instructed to press the '1' key on their keyboard to answer *Yes*, and the ‘2’ key to answer *No*. They were then presented with an empty grid and instructed to replicate the grid that they had seen at the start of that trial. They used the mouse to fill in the cells in the grid and to continue to the next trial.

The experiment started with three practice trials which matched the general format of the inference task but did not contain scalar expressions (the practice items are listed in Appendix A). After the practice trial, each participant was tested on all 24 items in Appendix A. The items were presented in random order.

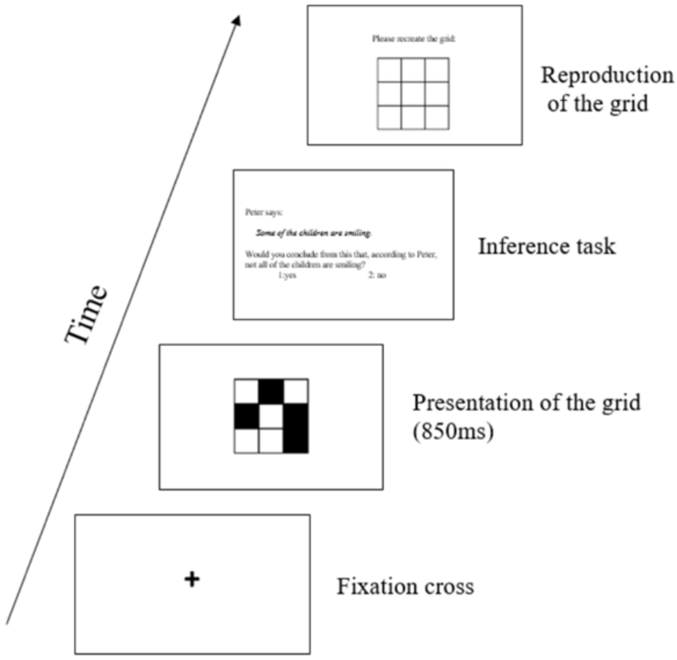

Figure 4 Depiction of the structure of a trial.

6.3 Data Treatment

Results from one participant failed to register. Two participants were excluded from the analysis because their accuracy on the memory task was under 75%. Accuracy was measured by dividing the number of correct squares by the total number of squares. 77 participants were therefore included in the final analysis (39 in the low load condition, 38 in the high load condition). Items with a response time below 200 ms and over 15 s were removed, as they are likely to be due to accidental keyboard responses or to a lapse in concentration.

6.4 Results

The data was analysed following the method used by Van Tiel et al. (2019). Model comparisons were conducted to test if a model with the effect of interest fits the data better than a maximally similar model without the effect. A significant main

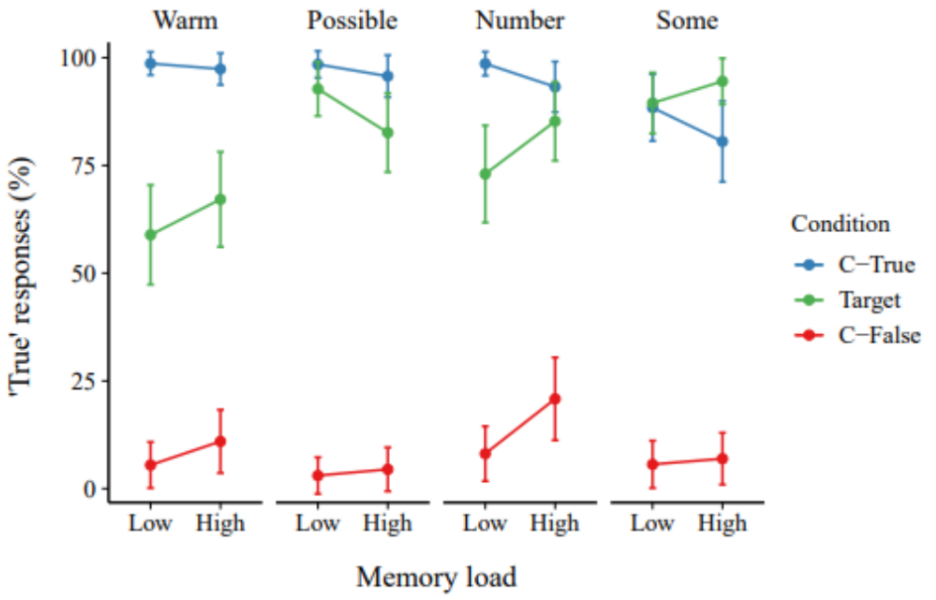

Figure 5 Percentage of "true" responses for each scalar term, condition and memory load. Error bars represent standard errors of the mean.

effect of scale type (x^2 (3)=48.83, $p < .001$) and an interaction of load and scale type (x^2 (3)=8.04, p = .045) were found.

To analyse effects of memory load, I constructed, for the target condition of each scalar word, a generalised mixed effects logistic regression model predicting responses ('yes' or 'no') on the basis of memory load (low load, or high load), including random intercepts for participants. Memory load was included as an ordinal factor. There was a marginal significant effect of memory load on the probability of 'true' responses for 'possible' (β = -0.70, SE = 0.40, Z = -1.76, p = .078). There were no effects of memory load for the remaining scalar words (all Z's < 1).

The main effect of scale type, the significant interaction between scale type and load, and the fact that ⟨possible, certain⟩ shows a marginally significant decrease in levels of yes responses in high load compared to low load fit our first prediction and show that the inference task can be used to test effects of memory load. No effect of memory load was found with ⟨some, all⟩, but it is worth noting that the effect found by Van Tiel et al. (2019) only existed between a no-load condition that was not included in our experiment and the low load, rather than between the low load and the high load.

For ⟨warm, hot⟩ and numbers, the graph in Figure 5 seems to fit the prediction of the proportion of *yes* ratings increasing under memory load. This effect, however, is not significant, which means that this observation cannot bring significant evidence to support that part of the predictions.

The last prediction, however, was that numbers would be processed like ⟨warm, hot⟩, and ⟨some, all⟩ would be processed like ⟨possible, certain⟩. To analyse effects of scale, I constructed, for the target condition of each scalar word, a generalised mixed effects logistic regression model predicting responses (*yes* or *no*) on the basis of scale (*some, possible, warm or number*), including random intercepts for participants. Scale type was included as an ordinal factor.

Under low load, there was no difference between numbers and ⟨warm, hot⟩ (β = 0.71, SE = 0.41, Z = 1.75, p = .080), or between ⟨possible, certain⟩ and ⟨some, all⟩ (β =-0.42, SE = 0.62, Z = -0.68, p = .495). Rates of *yes* responses were higher for ⟨possible, certain⟩ than ⟨warm, hot⟩ (β = -2.47, SE = 0.57, Z = -4.36, p $<$ .001), ⟨some, all⟩ than ⟨warm, hot⟩ (β = 2.05, SE = 0.50, Z = 4.18, p $<$ .001), ⟨possible, certain⟩ than numbers (β = 1.75, SE = 0.58, Z = 3.03, p = .002) and ⟨some, all⟩ than numbers (β =1.33, SE =0.50, Z = 2.62, p = .008). This fits the prediction that numbers are processed like ⟨warm, hot⟩ and ⟨some, all⟩ is processed like ⟨possible, certain⟩, and that ⟨warm, hot⟩ and numbers had lower rates of *yes* responses under low load than ⟨some, all⟩ and ⟨possible, certain⟩.

This paper did not make any predictions for the high load condition, apart perhaps from the prediction that there would be no effect at all, as half of the scales were increasing from a lower point and half were decreasing from a higher point, which could predict that they would end up at a similar middle point. Some effects of scale were found in the high load condition. Rates of yes responses for ⟨warm, hot⟩ were lower than those for ⟨possible, certain⟩ (β = 1.63, SE = 0.34, Z = 4.79, p $<$ .001), numbers (β = 1.89, SE = 0.31, Z = 2.91, p = .003) and ⟨some, all⟩ (β = 2.14, SE = 0.38, Z = 5.71, p $<$ .001). Rates of *yes* for ⟨possible, certain⟩ were lower than those for numbers (β =-0.73, SE = 0.36, Z = -2.01, p = .044), and the rate for numbers was lower than the rate for ⟨some, all⟩ (β = 1.25, SE = 0.40, Z = 3.14, p = .001). The theories presented in this paper do not bring a specific explanation for these results, but they are not problematic for our general findings. Further research could be done to explore them in more detail.

7 Discussion

First of all, the results show an effect of memory load for the ⟨possible, certain⟩ scale. This replicates previous findings using a working memory task, and adds to the theory that processing scalar inferences is cognitively difficult (Bott & Noveck 2004, Breheny et al. 2006, De Neys & Schaeken 2007). Indeed, participants prefer the reading of *possible* that is compatible with *certain* when the task's working memory demands increase. It therefore seems that the memory task is keeping them from making the *possible but not certain* inference, which means that that inference is in some way cognitively difficult. It is hard to know the nature of that processing difficulty, however, as previous research has shown that the observed effect could be due to the cognitive cost of processing the negative information in positively scalar items rather than the processing cost of the inference process itself (Van Tiel et al. 2019). The present results are not informative on this issue, as all of the chosen scales are positively scalar, so no comparison between positive and negative scales can be made. Further research could be done with negative scales using the inference task to shed more light on this issue.

This finding is also interesting for methodological reasons: indeed, the existence of such an effect shows that it is possible to use an inference task to find effects of memory load. This means that this same inference task could be used to test all of the items inVan Tiel et al. (2016), particularly those that cannot be represented visually, such as for example ⟨intelligent, brilliant⟩.

It is important to note, however, that this effect was only marginally significant, and no similar effect was found for the scales other than ⟨possible, certain⟩. A potential issue could be that the low load condition is already too cognitively demanding and impacts processing, which means that increasing the difficulty does not impact the responses significantly. It could therefore be interesting to compare this low load condition with the complete absence of any cognitive task. A potential follow-up to see a clearer effect would be to run the study with a third load condition in which participants only react to the inference task without the dual-task procedure involving a memory task. Indeed, in the Van Tiel et al. (2019) study, for the ⟨some, all⟩ scale the effect of load was between the no-load and low load conditions rather than the low load and high load conditions. Here there was an effect in the ⟨possible, certain⟩ scale, albeit a marginally significant one, which replicates results found for the *might* scale in Van Tiel et al. (2019). It is therefore likely that adding a no-load condition to the design would replicate the effects of Van Tiel et al. (2019) for ⟨some, all⟩, namely an effect between no-load and low load. If the addition of this condition makes it possible for an effect of load to appear, I would also expect to see a reverse effect of memory load in the adjective and number scales. Following the general pattern of the results, previous findings as well as the theoretical predictions laid out in this paper, my predictions for a no-load condition are the following: the rates of *yes* responses, namely the rates of interpreting the scalar in a lower-bounded way, should be lower in the no-load condition than in the low load condition for ⟨warm, hot⟩ and numbers, and higher in the no-load condition than in the low load condition for ⟨some, all⟩ and ⟨possible, certain⟩, which are the same predictions as those originally mapped out for the comparison between low load and high load.

For numbers, the graph and means follow the pattern found in Marty et al. (2013) for numbers: the average rate of *yes* responses is lower in low load than in high load. ⟨warm, hot⟩ follows the same pattern, which is what was predicted. There is no statistically significant effect of load within those scales, however the addition of the no-load condition as described above would be a good way to explore whether a significant effect can be found. Moreover, *yes* responses under low load are the same for numbers and *warm*, and lower than *yes* responses for ⟨some, all⟩ and ⟨possible, certain⟩, which matches the predictions mapped out in the introduction. Indeed, this shows a similarity in processing between numbers and adjectives such as *warm*, which was one of the theoretical questions that I set out to test. The no-load task could perhaps be useful to add more statistical significance to my findings, but the general pattern does follow the theoretical prediction set out in this paper. Indeed, the findings about the rates of ⟨warm, hot⟩ in the so-task led me to the theory that *warm* could have a primary exact meaning and be processed in a similar way to numbers, which would increase rates of *yes* in the inference task under memory load, and the data hints at that. It would be interesting to design further experiments to bring more support to this theory.

8 Conclusion

This paper set out to replicate previous results found on effects of memory load on scalar implicatures using a text inference task rather than a visual task. Results showed an effect of memory load with one of the scales, ⟨possible, certain⟩, which supports the theory of scalar inferences being cognitively difficult and creates a new potential methodology to test scalars, including expressions that cannot easily be visually represented.

The pattern for numbers and ⟨warm, hot⟩ also seems to be the opposite of the pattern for ⟨some, all⟩ and ⟨possible, certain⟩ which goes in the same direction as previous results found for numbers, and suggests that scalars such as ⟨warm, hot⟩ which rate high on the so-test could be processed in the same way as numbers, namely with a primary exact reading. Follow up experiments are necessary, however, as some of the patterns observed in the data did not reach statistical significance.

References

Bott, L. & I. A. Noveck. 2004. Some utterances and underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51(3). 437–457.

Breheny, R. 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25(2). 93–139.

Breheny, R., N. Katsos & J. Williams. 2006. Are generalised scalar implicatures generated by default? an on-line investigation into the role of context in generating pragmatic inferences. *Cognition* 100(3). 434–463.

Carston, R. 2011. Informativeness, relevance and scalar implicature. In R. Carson & S. Uchida (eds.), *Relevance Theory: Applications and implications* Pragmatics And Beyond New Series, 327–353. John Benjamins Publishing Company.

De Neys, W. & W. Schaeken. 2007. When people are more logical under cognitive load: Dual task impact on scalar implicatures. *Experimental psychology* 54(2). 128–133.

Geurts, B. 2006. The meaning and use of a number word. *Non-definiteness and plurality* 95. 311–329.

Grice, H. P. 1975. Logic and conversation. In *Speech acts*, 41–58. Leiden, Netherlands: Brill.

Horn, L. 1972. *On the semantics and pragmatics of logical operators in english*: University of California Los Angeles dissertation.

Levinson, S. 2000. *Presumptive meanings.* Cambridge, MA: MIT Press.

Marty, P., E. Chemla & B. Spector. 2013. Interpreting numerals and scalar items under memory load. *Lingua* 133. 152–163.

Moxey, L. M. 2006. Effects of what is expected on the focussing properties of quantifiers: A test of the presupposition-denial account. *Journal of Memory and Language* 55(3). 422–439.

Noveck, I. A. 2001. When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition* 78(2). 165–188.

Papafragou, A. & J. Musolino. 2003. Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition* 86(3). 253–282.

Sperber, D. & D. Wilson. 1986. *Relevance: Communication and cognition*, vol. 142. Cambridge, MA: Harvard University Press.

Sun, C., Y. Tian & R. Breheny. 2018. A Link Between Local Enrichment and Scalar Diversity. *Frontiers in Psychology* 9. 2092.

Van Tiel, B., E. Pankratz & C. Sun. 2019. Scales and scalarity: Processing scalar inferences. *Journal of Memory and Language* 105. 93–107.

Van Tiel, B., E. Van Miltenburg, N. Zevakhina & B. Geurts. 2016. Scalar diversity. *Journal of Semantics* 33(1). 137–175.

Blanche Gonzales de Linares
University of Cambridge
bmg39@cam.ac.uk

Appendix A

Practice items:

Mary says:

Peter lives in London.

Would you conclude from this that, according to Mary, *Peter lives in the United Kingdom?*

John says:

The table is dirty.

Would you conclude from this that, according to Mary, *the table is clean?*

Jane says:

Peter is short.

Would you conclude from this that, according to Jane, *Peter is blond?*

Some:

Target:

Peter says:

Some of the children are smiling.

Would you conclude from this that, according to Peter, *not all of the children are smiling?*

Jane says:
Some of the exams are difficult.
Would you conclude from this that, according to Jane, *not all of the exams are difficult?*

False control:
John says:
All of the students went to the party.
Would you conclude from this that, according to John, *some of the students did not go to the party?*

Mary says:
Some of the plants have flowers.
Would you conclude from this that, according to Mary, *none of the plants have flowers?*

True control:
Peter says:
All of the files have been deleted.
Would you conclude from this that, according to Peter, *at least some of the files have been deleted?*

Jane says:
All of the cows have been milked.
Would you conclude from this that, according to Jane, *at least some of the cows have been milked?*

Numbers:
Target:
John says:
Three horses are in the field.
Would you conclude from this that, according to John, *it is not true that four horses are in the field?*

Mary says:
Four chairs are in the room.
Would you conclude from this that, according to Mary, *it is not true that chairs are in the room?*

False control:
John says:
Three computers are broken.
Would you conclude from this that, according to John, *no more than two computers are broken?*

Jane says:

Four birds are flying.

Would you conclude from this that, according to Jane, *no more than three birds are flying?*

True control:

Mary says:

Four balls are bouncing.

Would you conclude from this that, according to Mary, *more than three balls are bouncing?*

Jane says:

Three girls are dancing.

Would you conclude from this that, according to Jane, *more than two girls are dancing?*

Certain/possible:

Target:

John says:

It is possible that Mary will come to the party.

Would you conclude from this that, according to John, *it is not certain that Mary will come to the party?*

Peter says:

It is possible that the train will arrive on time.

Would you conclude from this that, according to Peter, *it is not certain that the train will arrive on time?*

False control:

Peter says:

It is certain that it will rain tomorrow.

Would you conclude from this that, according to Mary, *it is not possible that it will rain tomorrow?*

John says:

It is certain that the cat will eat the mouse.

Would you conclude from this that, according to John, *it is not possible that the cat will eat the mouse?*

True control: Mary says:

It is certain that the shop will be open on Sunday.

Would you conclude from this that, according to Mary, *it is more than likely that the shop will be open on Sunday?*

Peter says:
It is certain that Mary will attend the lecture.
Would you conclude from this that, according to Peter, *it is more than likely that Mary will attend the lecture?*

Warm/hot:
Target:
John says:
The tea is warm.
Would you conclude from this that, according to John, *the tea is not hot?*

Peter says:
The weather is warm.
Would you conclude from this that, according to Peter, *the weather is not hot?*

False control:
John says:
The radiator is hot.
Would you conclude from this that, according to John, *the radiator is no more than warm?*

Mary says:
The soup is warm.
Would you conclude from this that, according to Mary, *the soup is cold?*

True control:
Jane says:
The water is hot.
Would you conclude from this that, according to Jane, *the water is more than warm?*

Mary says:
The bath is hot.
Would you conclude from this that, according to Mary, *the bath is more than warm?*