Scalar implicatures processing: slowly accepting the truth (literally)

Florian Schwarz — University of Pennsylvania
Jacopo Romoli — Ulster University
Cory Bill — CCD/Macquarie University

Abstract. The processing of scalar implicatures has been extensively studied in recent years (Bott and Noveck 2004; Bott et al. 2012; Breheny et al. 2006; Huang and Snedeker 2009, Chemla and Bott 2013 among others). Following the work of Bott and Noveck (2004), there has been an ongoing discussion about whether scalar implicatures are delayed in online processing. A second question, discussed more recently, is whether so-called 'indirect' scalar implicatures (e.g., *John didn't always go to the movies* $\rightsquigarrow$ *John sometimes went*) exhibit the same processing profile as the standard direct ones (e.g., *John sometimes went to the movies* $\rightsquigarrow$ *John didn't always go*). Cremers and Chemla (2014) argue that they do, whereas Romoli and Schwarz (2015) find the reverse pattern for indirect scalar implicatures, with slower Reaction Times associated with literal responses. In this paper, we report an experiment investigating two key questions arising from this debate: first, do Reaction Times yield uniform evidence for delayed availability of implicatures? and second, do direct and indirect scalar implicatures display comparable or distinct processing properties? Using the covered picture version of a picture matching task (Huang et al. 2013) with reaction time measures, we look both at cases where the overt picture is rejected and where it is accepted. Our results provide a negative answer to the first question: while delays for implicature responses arise for rejections (Covered picture choices), the pattern flips for acceptances (Target choices). The cross-over interaction in the results is inconsistent with attributing delayed implicature-rejections to delays in the availability of the implicature, since Target choices compatible with the implicature are faster than ones that are only compatible with a literal interpretation. This yields a very different perspective on reaction time results for implicatures. As for the second question, our results show that once acceptance vs. rejection is factored in, the general reaction time pattern is the same for indirect and direct scalar implicatures and compatible with the different results found in the literature.

Keywords: scalar implicatures, processing, Reaction Times

1. Introduction

Upon hearing a sentence like (1-a), we tend to conclude (1-b).[1] Similarly, and symmetrically, when somebody tells us (2-a) we typically infer (2-b). (1-b) and (2-b) are examples of 'scalar implicatures,' one of the most studied types of inferences that we draw from sentences.

[1]For stimulating discussions related to this work, we'd like to thank Lewis Bott, Emmanuel Chemla, Stephen Crain, Danny Fox, Dan Grodner, Napoleon Katsos, Lynda Kennedy, Clemens Mayr, Kelly Rombough, Raj Singh, Benjamin Spector,Yasutada Sudo, Rosalind Thornton, Lyn Tieu, and audiences at Sinn und Bedeutung 19, the Institute for Research in Cognitive Science at Penn, and Rutgers. Also thanks to Dorothy Ann for the pictures.

(1) a. John sometimes went to the movies.
b. ⇝*John didn't always go*

(2) a. John didn't always go to the movies.
b. ⇝*John sometimes went*

As we review in more detail below, one of the main characteristic of these inferences is that they can be suspended or cancelled. A theory of scalar implicatures therefore has to account both for how scalar implicatures arise when they do arise, and for the circumstances in which they do not arise. While this is quite generally agreed upon, the question of how exactly scalar implicatures should be derived is controversial, and a variety of theories have been proposed, which differ substantially in the mechanisms behind these inferences. It is also controversial where exactly these inferences come into play at the interface between semantics and pragmatics.[2] Crucially for our purposes, however, there is a solid consensus on treating the inferences in (1-b) and (2-b) uniformly as the same type of inference. While they sometimes are distinguished terminologically, e.g., by calling (1-b) 'direct' scalar implicatures and (2-b) 'indirect' ones (Chierchia 2004), this is not usually assumed to involve a difference in the mechanisms that give rise to them.

More recently, scalar implicatures have also been extensively studied in the psycholinguistic literature, where they have played a central role in efforts to try to understand how theoretical distinctions between aspects of meaning map onto actual cognitive processes in language comprehension. The focus of this work has been to understand how these inferences are processed, in particular concerning the time course of processing implicatures in comparison to 'literal' content (Bott and Noveck 2004; Bott et al. 2012; Breheny et al. 2006; Huang and Snedeker 2009, Chemla and Bott 2013; see also Chemla and Singh 2014 for a critical review). Following the seminal work of Bott and Noveck (2004), there has been an ongoing discussion about whether scalar implicatures are delayed relative to literal content in online processing. Bott and Noveck (2004) provided Reaction Time data they argued to support the existence of a delay. Various later variations of their study point in the same direction (see, for example, Bott et al. 2012, who look at speed accuracy trade-offs and Chemla and Bott 2013, who compare scalar implicatures to so called 'free choice' inferences), and Breheny et al. (2006) present self-paced reading they argue to support this interpretation as well. More recently, the issue has been investigated using the visual world paradigm, where results have been more mixed. Huang and Snedeker (2009, and subsequent work) report delays in eye movements based on the 'not all' implicature of 'some'. But various others, e.g., Grodner et al. (2010), Breheny et al. (2013), and Degen and Tanenhaus (2011), report eyetracking results which they argue show that scalar implicatures are available immediately, i.e., with no delay. There thus is a contrast in the overall findings between methodologies, as – to our knowledge – reaction time studies have consistently found delays for implicature-based responses. Such apparent differences across methodologies should be accounted for if they persist, and the studies presented here provide a new perspective on reaction time data.

[2] See Chierchia et al. 2012; Sauerland 2004, 2012; Geurts 2010; Schlenker 2012 among others.

Another question concerning the processing of scalar implicatures, which has only been investigated more recently, is whether indirect scalar implicatures (ISIs) exhibit the same processing profile as the standard direct ones (DSIs). Cremers and Chemla (2014) argue that they indeed yield comparable results, with faster Reaction Times (RTs) for literal responses. Romoli and Schwarz (2015), however, find the reverse pattern for indirect scalar implicatures, with **slower** RTs for literal responses. In this paper, we report on an experiment investigating two key questions arising from this debate:

(3) a. **Question 1:** Do RTs yield uniform evidence for delayed availability of implicatures?
b. **Question 2**: Do DSIs and ISIs display comparable or distinct processing properties?

Using the covered picture (a.k.a. Covered Box) version of a picture matching task with RTs (Huang et al., 2013), we look at two different types of responses relating to implicatures. Firstly, there are cases where an overtly shown picture is rejected based on an implicature; secondly, there are cases where an overt picture that is compatible with an implicature interpretation is accepted. The previous literature has focused on the former, in the form of 'false'-judgments in truth-value judgment tasks, but it turns out that looking at both types of cases enhances our perspective on the time course of implicature processing substantially. To foreshadow, our results suggest a negative answer to **Question 1**: while delays for implicature responses arise when we look at rejections (Covered picture choices), the pattern flips when we look at acceptances (Target choices). The cross-over interaction in the results is inconsistent with an interpretation that attributes delayed implicature-rejections to delays in the availability of the implicature, since Target choices compatible with the implicature are **faster** than ones that are only compatible with a literal interpretation. This yields a very different perspective on reaction time results for implicatures. As for **Question 2**, our results show that once the importance of distinguishing acceptance and rejection responses is acknowledged, the general RT pattern for ISIs and DSIs is the same and compatible with both of the seemingly inconsistent results from the recent literature (Cremers and Chemla, 2014; Romoli and Schwarz, 2015).

The paper is organized as follows: in section 2, we briefly review the main characteristics of DSIs and ISIs and sketch their uniform derivation. In section 3, we discuss results from previous reaction time studies on scalar implicatures, in particular focusing on the recent contrasting results on indirect scalar implicatures mentioned above. In section 4, we report a new experiment investigating the questions in (3) and discuss the implication of its results. Section 5 concludes.

2. Direct and indirect scalar implicatures: characteristics and derivation

We begin by briefly reviewing the main properties of the phenomena under investigation, and sketch the uniform derivation of direct and indirect scalar implicatures that is standard in the literature. As only very general properties of these accounts will be relevant for our purposes, this review will remain fairly superficial.[3] As discussed above, a sentence like (4-a) gives rise to the inference in (4-b). In parallel, a sentence like (5-a) comes with the inference in (5-b).

(4) a. John sometimes went to the movies.
b. $\rightsquigarrow$*John didn't always go*

(5) a. John didn't always go to the movies.
b. $\rightsquigarrow$*John sometimes went*

These inferences are not obligatorily present, but can be suspended or cancelled. For instance, (4-a) can be followed felicitously by a continuation that is incompatible with the inference (4-b).

(6) John sometimes went to the movies. In fact, he always went!

In this regard, implicatures differ from literal, entailed content, as witnessed by the contrast between (4-b)/(5-b) and (7-b), which is a simple entailment of (7-a).[4] There is no way of canceling or suspending the inference in (7-b) without sounding contradictory, as illustrated in (8).

(7) a. John and Mary went to the movies.
b. $\rightsquigarrow$*John went to the movies*

(8) John and Mary went to the movies. #In fact, John didn't go!

A theory of scalar implicatures thus has to explain both the presence and absence of the relevant inferences in the various circumstances. The general idea behind virtually all accounts of scalar implicatures is that they arise from a comparison of the utterance made with certain alternative utterances; the latter are then inferred to be false, given certain assumptions about the speaker behaving rationally in communication.[5] In the cases at hand, the reasoning is as follows:

[3]The reader interested in more information on scalar implicatures should consult Schlenker 2012 and references therein.

[4]Where entailment is standardly defined as in (i):

(i) For any sentence p, q and possible world w, p ENTAILS q iff the following hold:
if p is true in w then q is true in w.

[5]This idea of implicatures arising from a comparison between what the speaker said and what she might have said instead goes back to the work of Grice 1975.

(9) For any sentence A:
 a. The speaker uttered A
 b. Alternatively, the speaker could have uttered the more informative sentence B.
 c. The most plausible reason for not uttering B (given some additional assumptions) is that B is false (or at least taken to be false by the speaker).

One crucial question is just what set of alternative utterances B should be considered. The general answer in the literature, since Horn 1972, is that certain words like *some* and *all*, or *sometimes* and *always* are associated with one another through a scale based on entailment.[6] Sentences containing one of these items then can be associated with a set of alternative sentences where the original item has been replaced by alternatives from its scale. For instance, for a sentence like (10-a), the sentence in (10-b) is an alternative that results from replacing *sometimes* with *always*.

(10) a. John **sometimes** went to the movies.
 b. John **always** went to the movies.

Following the schema in (9), the sentence in (10-a) is compared to that in (10-b), and the hearer will reason that while the speaker said the former, she could have said the latter instead. Since (10-b) is stronger than (10-a), the hearer will conclude that (10-b) is false, i.e., infer (11).

(11) John didn't always go to the movies.

The exact same reasoning applies to (5-a): the speaker uttered (5-a) but also must have considered (the equivalent of) (12). (12) is constructed by replacing *always* with its associates *sometimes* and which in this case is stronger than (5-a) due to negation. Again, the hearer concludes that the stronger sentence is false, i.e., infers that (12) is false. But the negation of (12) is of course equivalent to the inference in (5-b).

(12) John didn't **sometimes** go to the movies.

All theories that we know of provide a scalar implicature algorithm based on alternatives of this sort and derive direct and indirect scalar implicatures in a unified fashion (see Chierchia et al. 2012 for discussion). In the next section, we turn to the literature on processing of scalar implicatures, focusing in particular on the comparison of the processing of direct and indirect scalar implicatures.

[6]For more recent and more developed theories of alternatives see Katzir 2007 and Fox and Katzir 2011.

3. The processing of direct and indirect scalar implicatures

In recent years, research on scalar implicatures underwent what Chemla and Singh (2014) call an 'experimental turn.' In particular, investigations of their processing properties have played a central role in the overall theoretical discussion. Most studies have focused on direct scalar implicatures, but some recent studies have started looking at indirect ones too. In this section, we first review the classical findings on the processing of direct scalar implicatures, and then discuss two contrasting results on indirect scalar implicatures, which in large part motivate the new experiment reported below. We primarily focus on RT studies here, but eventually the goal is to integrate results obtained with other methodologies into a more comprehensive perspective. We will come back to this briefly in the final section.

3.1. The classical finding on direct scalar implicatures

Bott and Noveck (2004) found direct scalar implicatures to be associated with a delay in RTs. To sketch the gist of their result, they used sentences like (13-a), which gives rise to the direct scalar implicature in (13-b), which in turn directly conflicts with the relevant piece of common knowledge (i.e., that all elephants are mammals). Therefore, the sentence in (13-a) should be judged 'false,' under an inference interpretation (i.e., an interpretation that includes the inference).[7] As we know, however, the sentence can also be understood in a way that does not include the inference, and this reading is compatible with common knowledge (i.e., the literal meaning is equivalent to 'some or all elephants are mammals'). Under such a reading, which we will refer to as a 'no-inference reading', (13-a) should then be judged 'true.'

(13) a. Some elephants are mammals.
b. $\rightsquigarrow$ *Not all elephants are mammals*

The logic of Bott & Noveck's design then is as follows: since 'false' responses are indicative of inference interpretations and 'true' responses of no-inference interpretations, measuring RTs for both types of responses should shed light on the time course of the availability of the two interpretations.[8] Their main finding is that false responses were slower than true ones, and they interpret this as showing that scalar implicatures are associated with a delay relative to literal meaning. Schematically, we can summarize their result as: inference $>$ no-inference .[9]

[7] Notice that the sentence in (13-a) is generally judged to be an odd sentence, like all sentences which give rise to scalar implicatures conflicting with common knowledge (Magri 2010). This feature of the design is however shown not to be important in work replicating the main result of Bott and Noveck (2004), like that of Chemla and Bott (2013).

[8] There is an obvious potential concern about general difference between the time course of true and false responses, which Bott & Noveck try to address through various variants of their basic design. We will return to this issue when introducing our own study below.

[9] Where $\phi > \psi$ is supposed to be read as 'the RT associated with ϕ is larger than that associated with ψ.'

One particularly relevant version of their general approach is based on training participants to respond according to one or the other possible interpretations of the sentence in question. They find that participants that were trained to respond based on the no-inference interpretation were generally faster than those trained on the inference interpretation. Parallel results have been obtained in various similar studies since (Chemla and Bott 2013; Bott et al. 2011 among others), and also for implicatures with disjunction (Chevallier et al. 2008). As already mentioned, other methodologies, such as reading times (Breheny et al. 2006) and visual world eye tracking (Huang and Snedeker 2009) have yielded comparable results as well, though some researchers have argued results from the latter method to show that implicatures are immediately available.

3.2. Contrasting results on indirect scalar implicatures

Cremers and Chemla 2014 Cremers and Chemla (2014) extend Bott and Noveck's approach to indirect scalar implicatures by looking at sentences like (14-a), with the inference in (14-b), again incompatible with common knowledge.

(14) a. Not all elephants are reptiles.
b. $\rightsquigarrow$ *Some elephants are reptiles*

Overall, they argue that their findings are parallel to Bott and Noveck's results, in that training participants to respond based on an inference interpretation vs. a no-inference interpretation gives rise to the result that the inference-readings were slower than no-inference ones. That is, the inference group, responding 'false' to (14-a) was slower on average than the no-inference group, responding 'true.' Again, schematically, we could summarize their finding as: inference > no-inference.[10]

In sum, Bott and Noveck found that 'false' responses to sentences whose direct scalar implicature was false were slower than 'true' responses based on no-inference interpretations. Similarly, Cremers and Chemla found that 'false' responses to sentences whose indirect scalar implicature was false were slower in comparison to 'true' responses. These results are entirely in line with the general uniformity for direct and indirect scalar implicatures assumed in the literature, and with the initial interpretation by Bott and Noveck that scalar implicatures are associated with a delay.

Romoli and Schwarz 2015 Another recent study that includes reaction time results for indirect scalar implicatures is reported in Romoli and Schwarz (2015). The more general context was

[10] Cremers and Chemla (2014) report two experiments. In the first one, without training, they actually found opposite results for DSIs and ISIs. That is, they found that participants were faster to answer False than True for ISIs. They argue that this outcome is the result of a confound introduced by the presence of negation.

Always-Target (Inference) *Always*-Target (No-Inference)

Figure 1: Target pictures for *always* conditions from Romoli and Schwarz (2015).

a comparison between presuppositions and scalar implicatures under negation.[11] Rather than a direct truth-value judgment task, Romoli and Schwarz (2015) used a version of a sentence picture matching task with a covered picture (Huang et al. 2013), with both response choices and reaction time measures as dependent variables. Participants were told to select the best match for a given sentence from a set of three pictures (depicting individuals and a calendar strip with various iconic representations of activities for each day), while assuming that a) the 'covered picture' represented a hidden picture and b) that only one of the three pictures did indeed match the sentence. The overt pictures included a target picture that either was or was not consistent with the inference ('Inference' versus 'No-Inference' condition). The second overt picture served as a distractor. If neither of the overt pictures was considered a match, participants were instructed to click on the covered picture.

In contrast to previous studies, where different interpretations were associated with different types of responses, this setup made it possible to compare alike responses - namely acceptance responses in terms of target picture choices - based on different interpretations. For example, participants would see a sentence like (15), and evaluate it relative to target pictures such as those in Figure 1.

(15) John didn't always go to the movies last week.

The inference picture, where the depicted character indeed went to the movies on several days, is consistent with the 'sometimes' implicature of 'not always', whereas the no-inference picture is not. Looking at RTs for target choices in both conditions thus allows for a comparison of different interpretations giving rise to the same type of response, in terms of target choices.[12] The results revealed RTs in the Inference condition to be significantly faster than in the No-Inference condi-

[11] An example of a presupposition is in (i-b) from (i-a).

(i) a. John didn't stop going to the movies on Wednesday.
 b. ⇝ *John used to go to the movies before Wednesday*

[12] In principle, the inference picture could also be chosen based on a literal interpretation. But if all such choices were based on a literal interpretation, we would not expect any differences between the two conditions, so given the observed reaction time differences, at least a sizable portion of responses seems to result from inference interpretations.

tion, which directly contrasts with previous findings. Somewhat less surprisingly, target choices were much more frequent in the Inference condition.[13] Schematically, the result from Romoli and Schwarz is: inference $<$ no-inference.

These results are puzzling from the traditional perspective that ISIs are generated in the same way as DSIs and appear to be exactly the opposite of what Cremers and Chemla (2014) find: the literal (no-inference) responses in Romoli and Schwarz (2015) for indirect scalar implicatures were slower than implicature (inference) responses. The strong bias in the target-choice frequency for inference-based interpretations also is somewhat at odds with previous findings in comparable task settings for direct implicatures, where literal interpretations tend to be quite freely available. The question that arises is therefore what is behind these seemingly conflicting findings. They could be due to general task-related effects of various sorts. But a more particular difference between the studies concerns the types of responses that are compared. Previous reaction time studies quite generally wound up comparing 'true' responses to 'false' responses, whereas the setup of Romoli and Schwarz (2015) compared acceptance responses based on different interpretations. This raises the possibility that looking at different responses for different interpretations introduces a confound, either based on general differences between 'true' and 'false' responses, or based on differential effects for inference-based responses for acceptance and rejection responses. The experiment reported in the next section was designed to shed light on this issue.

4. Experiment

The present experiment extended the approach taken in Romoli and Schwarz (2015) to assess whether there indeed are processing differences between direct and indirect implicatures when we compare alike responses, as in the previous study. In order to gain a more comprehensive perspective on the impact of the covered picture methodology on the result patterns, we did not only look at target choices (acceptance judgments) based on inference and no-inference interpretations, but also at covered picture choices (rejection judgments).

4.1. Methods

Materials & Procedure We once again used the Covered picture paradigm, with both response choices and RTs as dependent variables. To simplify the task, participants were only presented with two pictures, one of which was simply black and was described as covering a hidden picture. The instructions provided a detective scenario, where information about a suspect was presented as having been extracted from intercepted communication, and the participant's task was to decide

[13]The effects were parallel for the presupposition conditions looking at *stop*, where the effect was numerically even more pronounced, although there was no significant interaction in RTs between inference types. We refer the interested reader to Romoli and Schwarz (2015) for further details on *stop* as well as the larger theoretical relevance of the comparison between *always* and *stop*.

which of two potential culprits fit the provided description. It was explicitly stated that only one of the two pictures would match the description, so that the covered picture should be chosen in case the overt picture did not match the sentence. This setup was chosen to increase the chances of participants basing their responses on no-inference interpretations. First, the described source of the information remained opaque due to its nature of stemming from intercepted communication, which makes it uncertain whether the speaker of that sentence was fully informed. Secondly, the emphasis that only one picture would match the description provided by the sentence should increase target choices for no-inference pictures, assuming no-inference interpretations are in principle available but generally somewhat dispreferred.

The basic logic of the design was parallel to that of Romoli and Schwarz (2015), in that the overt target picture either was consistent with a given interpretation or not. More concretely, the sentences (a) and (b) in Figure 2 were displayed with one of the pictures in Figure 2 and a covered picture. Conditions were coded as indicated below the pictures.

a. b. c.

Inference-False	Inference-True	Literal-False	**(DSI)**
Literal-False	Inference-True	Inference-False	**(ISI)**

(a) John sometimes went to the movies ⇝ *John didn't always go* **(DSI)**

(b) John didn't always go to the movies ⇝ *John sometimes went* **(ISI)**

Figure 2: Target Picture versions and conditions

For the direct scalar implicature condition with *sometimes*, the picture in Figure 2a tested for literal, no-inference interpretations, as the depicted person always went to the movies. Target choices thus must be based on the no-inference interpretation. Covered picture choices for this picture in turn are indicative of inference interpretations. The picture in Figure 2b is consistent with an inference interpretation (as well as a no-inference interpretation, since it is entailed by the inference interpretation), so target choices are generally expected here. Finally, the picture in Figure 2c is inconsistent with both interpretations, as the depicted individual never went to the movies. For purposes of analysis, this design allowed us to compare both target and covered picture responses to the picture in Figure 2a to target and covered picture responses in the control conditions in Figures 2b and 2c respectively. We thus get a comparison between inference-based rejections (covered picture choices for Figure 2a) and literal meaning based rejections (covered picture choices for Figure 2c), as well as between no-inference acceptances (target choices for Figure 2a) and inference acceptances (target choices for Figure 2b). Figure 3 is a summary of the two comparisons: no-inference

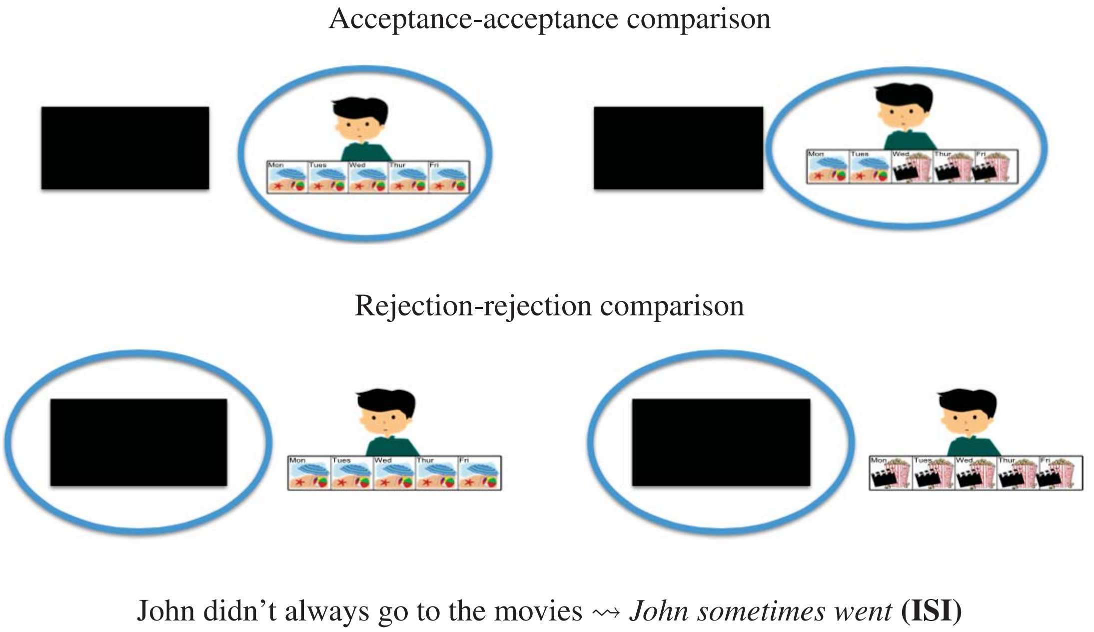

Figure 3: Acceptance-acceptance and rejection-rejection comparisons for ISI sentences

acceptance vs inference acceptance ('acceptance-acceptance' comparison) and inference-rejection versus no-inference rejection ('rejection-rejection' comparison).

Note that target choices for Figure 2b and covered picture choices for Figure 2c could in principle be based on either inference or no-inference interpretations, given the entailment relations between the two. However, to the extent that we find behavioral differences between these conditions and the relevant no-inference comparison conditions, we assume these to be due to a substantial proportion of inference interpretation-based responses in the former conditions.

The same logic applies to the ISI sentences, with different mappings onto the pictures. The picture in Figure 2c serves as a test for no-inference interpretations, as target choices are incompatible with the inference that John sometimes went to the movies. Covered picture choices for these pictures in turn must be based on inference interpretations. The picture in Figure 2b is consistent with the inference interpretation (as well as a no-inference interpretation, as for DSIs), and the picture in Figure 2a is inconsistent with either interpretation. So in the case of ISIs, Figure 2c is expected to yield a mix of target and covered picture choices, depending on the interpretation participants base their judgments on in a given trial, which can be compared to the covered picture and target choices in the respective control conditions.

Inference Type	Inference False (Figure 2a/c)	Literal False (Fig 2c/a)	Inference True (Fig 2b)
DSI	22.9	0.005	97.1
ISI	50.9	0.005	95.7

Table 1: Target choice rates in % by condition

Participants & Procedure 35 undergraduate students from Macquarie University participated in the study. They saw 36 sentence picture pairs of the sort described above, with 6 items for each pairing, counterbalanced across participant groups. In addition, there was a total of 36 filler items; 18 were variants of the experimental items containing *always* without negation, paired with all three picture types to ensure that pictures such as those in Figures 2a/c were viable target choices throughout the experiment sufficiently often. There also were 6 items containing plain negation (e.g., *John didn't go to the movies last week.*), again paired with the various picture types to even out choices of types of pictures. Finally, 12 items were from another sub-experiment containing negation and *again*. At the beginning of the experiment, participants were presented with instructions laying out the detective scenario described above. They then were shown some example sentences and pictures, and did a total of 4 practice trials (none of them resembling the crucial experimental conditions) to ensure they understood the covered picture setup. Throughout this initial phase, they were free to ask any clarification questions. After this, presentation of the experimental trials began.

4.2. Results

Responses were coded according to whether they were based on/compatible with literal content alone (1a/2c Target response; 1c/2a-Covered Picture or based on/compatible with the implicature (1a/2c-Covered Picture; 1b/2b-Target). Target choice proportions as well as RTs (measured from the display of the sentence, which was added to the screen 800ms after the picture was first shown) were analyzed.

Response Rates Mean target selection rates are provided in Table 1. Accuracy in the conditions where literal and implicature interpretations are consistent (Figures 2b/c for DSIs, Figures 2a/b for ISIs) were at ceiling, as expected. Both inference and no-inference interpretations occurred with both DSIs and ISIs, but inference interpretations occurred more often with DSIs than with ISIs, as there were fewer target choices for the Inference False picture for DSIs. A planned comparison between these two conditions using a logistic regression mixed-effect model revealed this difference in implicature-response rates to be significant ($\beta = 4.01$, $SE = 0.98$, $z = 4.07$, $p < .001$).

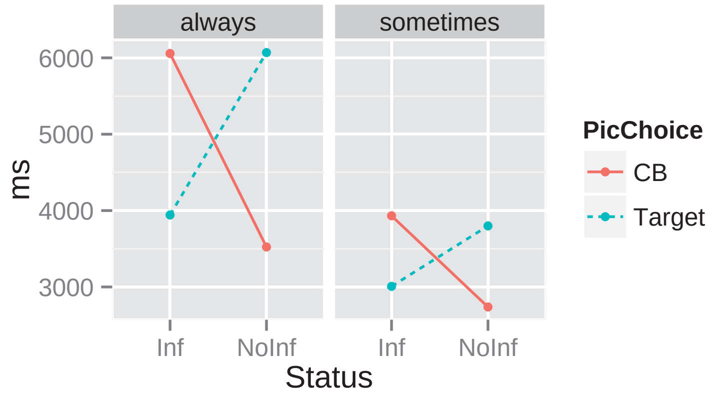

Figure 4: RTs for inference and no-inference responses by response type and implicature type.

Reaction Times For the purposes of analyzing response times, the trials in the Inference False condition were again categorized as inference and no-inference interpretations based on the response provided. Accurate responses in the Inference True and Literal False conditions were coded as inference and no-inference conditions respectively. The mean RTs for all conditions encoded in this manner are illustrated in Figure 4. As is apparent right away from the cross-over interaction pattern in the graph, the relation between RTs for inference and no-inference interpretations depends crucially on whether we look at acceptances in the form of target choices or rejections in the form of covered picture choices. In the former case, inference interpretations are faster than no-inference ones, while the reverse holds in the latter.

To investigate this result statistically, we analysed both the DSI and ISI subsets of data as a 2×2 interaction design, using mixed-effect models with subjects and items as random effects, as implemented in the *lmer* function of the *lme4* package in R (Bates 2005). Following Barr et al. (2013), we used the maximal random effect structure that would converge, with random effect slopes for each factor, as well as the interaction. To assess whether inclusion of a given factor significantly improved the fit of the overall model, likelihood-ratio tests were performed that compared two minimally different models, one with the fixed effects factor in question and one without, while keeping the random effects structure identical (Barr et al. 2013). We report estimates, standard errors, and t-values for all models, as well as the χ^2 and p-value from the likelihood-ratio test for individual factors. The statistical details are summarized in Table 2. The 2×2 interactions were highly significant for both ISIs and DSIs, as were the relevant simple effects comparing inference vs. no-inference responses by response type. Thus, schematically, what we found is in (16): inference-based rejections (Covered Picture-choices) were slower for both types of implicatures, in line with previous findings for DSIs from Bott and Noveck (2004) on, and with the findings by Cremers and Chemla (2014) for ISIs. However, looking at acceptances (Target-choices), those compatible with the implicature were faster than those only compatible with the literal meaning.

DSI's	β	SE	t	χ^2	p
Interaction	2119.1	563.4	3.76	9.67	$<$**.01**
Simple Effects:					
Covered Picture Choices: Inf > NoInf	-1418.6	534.8	-2.65	6.38	$<$**.05**
Target Choices: Inf < NoInf	666.1	276.5	2.41	5.42	$<$**.05**
ISI's					
Interaction	5902.7	1793.5	3.29	9.67	$<$**.01**
Simple Effects:					
Covered Picture Choices: Inf > NoInf	-3.302.2	881.6	-3.75	7.80	$<$**.01**
Target Choices: Inf < NoInf	2197.9	580.2	3.788	11.734	$<$**.001**

Table 2: Summary of reaction time analyses: Interaction between Response Choice and inference status and simple effects for relevant paired factor levels.

This is in line with the findings by Romoli and Schwarz (2015) for ISIs, but a novel finding for DSIs.

(16) RT patterns for Scalar Implicatures (for both DSIs and ISIs):
  a. **rejection response**
     Inference > no-inference
  b. **acceptance response**
     Inference < no-inference

4.3. Discussion

Our experimental investigation aimed to answer the two questions in (17), repeated from above:

(17) a. **Question 1**: Do RTs yield uniform evidence for delayed availability of implicatures?
  b. **Question 2**: Do DSIs and ISIs display comparable or distinct processing properties?

Our results suggest a negative answer to **Question 1**: While response times for rejection responses in the form of covered picture choices yielded the familiar pattern in line with the results from Bott and Noveck (2004) and following work on DSIs and from Cremers and Chemla (2014) on ISIs – with slower response times for inference responses than no-inference responses – response times for acceptance responses in the form of target choices gave rise to the exact opposite of that pattern, in that inference responses were **faster** than no-inference responses. The predominant interpretation of the common result for inference-based rejections in the literature is that the calculation of implicatures involves time-consuming processing efforts, which lead to a delay in the availability of an inference interpretation. However, this fails to account for the fact that acceptance judgments are faster for pictures consistent with the inference than for ones that are not. Since both are consistent with the no-inference interpretation, which is assumed to be available more or less immediately, this type of account would expect them to be be on par in terms of response times. Note furthermore that our results reconcile the apparent tension between the two previous studies on ISIs. The study in Romoli and Schwarz (2015) looked at acceptance responses in terms of target choices, and the pattern in the present results is entirely parallel to the finding reported there. At the same time, the pattern for rejection responses is parallel to that reported by Cremers and Chemla (2014). On a descriptive level, the present results thus validate both of the previous results, and contribute a new level of understanding to the role of the type of response for reaction time outcomes. As for **Question 2**, our results show that once acceptance/rejection is factored in, DSIs and ISIs yield the same pattern of RTs.

Turning to the interpretation of the present results, the question is how to account for the reverse patterns for acceptance and rejection judgments in our data. With regards to the former, the speed-up in the acceptance of implicature-compatible cases suggests that inference interpretations are available just as readily as no-inference interpretations, as pictures compatible with them are quickly identified as matches. But why are pictures that are only compatible with a literal interpretation accepted more slowly? This cannot be due to a delay in availability of the literal interpretation since a), the inference interpretation entails the literal interpretation and b), rejections of pictures based on the literal content are fast. Thus, an alternative explanation of both the acceptance and rejection patterns is called for. Our proposal for interpreting the present results starts from the observation that across all conditions, delays arise precisely in those circumstances where the two interpretations of the sentences that are in principle possible – what we have labeled the inference and no-inference interpretation – conflict with one another. For example, we find relatively slow target choice responses when the target is compatible with the no-inference interpretation but incompatible with the no-inference interpretation (Figure 2a for DSIs and Figure 2c for ISIs). Similarly, covered picture responses are also slow in the very same circumstances. The basic idea then is that there are opposing pressures favoring the respective interpretations, and that delays arise precisely when there is a conflict between such factors that affect the overall response. More specifically, we assume that comprehenders are generally charitable, i.e., they generally try to construe utterances in such a way that they are true of the circumstances at hand. In our case, charity can plausibly be seen as corresponding to selecting the overt picture, as that is the obvious and salient option at hand. On the other hand, it is intuitively plausible that inference interpretations for scalar implicatures are generally preferred. As anyone that has taught introductory logic can confirm, it takes some serious effort to convince students that *some*-statements are in principle compatible with universal scenarios, i.e., that *some* does NOT literally mean *some but not all*. The pressures of selecting the overt picture and the preference for scalar meanings oppose one another in precisely those conditions where we find a delay in our data. In the Literal False conditions (Figure 2a for DSIs and Fig 2c for ISIs, the principle of charity favors the Target, and the preference for scalar meanings favors the covered picture. Whether participants end up choosing the Target or the covered picture, their responses are delayed in these cases, compared to rejections (Figure 2c for DSIs and Figure 2a for ISIs) and acceptances (Figure 2b for both DSIs and ISIs) of the relevant controls. This account is similar in spirit to that by Katsos and Bishop (2011), who explain acquisition data in terms of pragmatic tolerance: from our perspective, this corresponds to claiming that the charity principle is stronger in children than the preference for implicature interpretations. Returning to the comparison between ISIs and DSIs, it is interesting that while as far as RTs are concerned, our results are comparable for ISIs and DSIs, the rate of implicature interpretations is significantly higher for DSIs. It is possible that this is simply due to complexities introduced by negation, but a more detailed explanation will have to be fleshed out in future work.

5. Conclusion

In this paper, we reported on the results of an experiment comparing DSIs and ISIs, looking at acceptance and rejection responses. For both DSIs and ISIs, we found that in the case of rejection, the RTs of inference responses were slower than that of no-inference ones. However, in the case of acceptance, we found the RTs associated with inference responses to be faster than those for no-inference ones. As discussed, these results are in line with the predicted uniformity of DSIs and ISIs and reconcile the apparently conflicting results in the literature on ISIs by Cremers and Chemla (2014) and Romoli and Schwarz (2015). Moreover, they challenge an interpretation of the results based on the rejection response pattern in (16-a) in terms of a general delay associated with the computation of implicatures. Such an interpretation leaves the RT pattern for acceptance responses in (16-b) unaccounted for. We proposed a different explanation encompassing both results based on two pragmatic principles, which lead to a delay when they end up conflicting with each other.

The present discussion focused on RT data. In integrating these with the more general results on scalar implicatures processing in the literature, one important question is how these RT results relate to results obtained with other methodologies. While we argue our results show that RT data for implicatures cannot be accounted for in terms of a delayed availability of implicatures, they leave open whether delays in early stages of processing as indicated by more immediate online measures do support such a notion. But at least some of the alleged evidence for delayed implicature computation has to be reconsidered in light of the present data. In ongoing work, we are extending the Covered picture paradigm to eyetracking in the visual world paradigm to collect more direct online measures on the time course of acceptance, with the goal of gaining a more comprehensive perspective on scalar implicature processing based on these different methodological approaches.[14]

[14]In other related work, a behavioral acquisition study on 4/5 year-old children and adults found indirect and direct scalar implicatures to differ, see Bill et al. (to appear). Also see Kennedy et al. (to appear) for an extension of this study to a population with Broca's Aphasia, which also exhibit different responses for the two types of implicatures. In ongoing work, we are now extending the investigation of the behavior of these populations with scalar implicatures on processing experiments.

References

Barr, D. J., R. Levy, C. Scheepers, and H. J. Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language 68*(3), 255–278.

Bates, D. M. (2005). Fitting linear mixed models in r. *R News 5*, 27–30.

Bill, C., J. Romoli, and F. Schwarz (to appear). Scalar implicatures vs presuppositions: the view from acquisition. *TOPOI*.

Bott, L., T. M. Bailey, and D. Grodner (2011). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language 66*(1), 123–142.

Bott, L., T. M. Bailey, and D. Grodner (2012). Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language 66*(1), 123–142.

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

Breheny, R., H. J. Ferguson, and N. Katsos (2013). Investigating the timecourse of accessing conversational implicatures during incremental sentence interpretation. *Language and Cognitive Processes 28*(4), 443–467.

Breheny, R., N. Katsos, and J. Williams (2006, July). 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.

Chemla, E. and L. Bott (2013). Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes 38*(3), 241–260.

Chemla, E. and R. Singh (2014). Remarks on the experimental turn in the study of scalar implicatures (part i and ii). *Language and Linguistics Compass 8*(9), 373–386.

Chevallier, C., I. Noveck, T. Nazir, L. Bott, V. Lanzetti, and D. Sperber (2008). Making disjunctions exclusive. *Quarterly Journal of Experimental Psychology 61*, 1750–1761.

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

Chierchia, G., D. Fox, and B. Spector (2012). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In C. Maienborn, K. von Heusinger, and P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning volume 3*, pp. 2297–2331. Berlin: Mouton de Gruyter.

Cremers, A. and E. Chemla (2014). Direct and indirect scalar implicatures share the same processing signature. In S. Pistoia Reda (Ed.), *Pragmatics, Semantics and the case of scalar implicatures*, Language and Cognition. Palgrave Macmillan. Ms., Ecole Normal Superieure, Paris.

Degen, J. and M. K. Tanenhaus (2011). Making inferences: The case of scalar implicature processing. In L. Carlson, C. Hölscher, and T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pp. 3299 – 3304.

Fox, D. and R. Katzir (2011). On the characterization of alternatives. *Natural Language Semantics 19*(1), 87–107.

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

Grice, H. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and semantics: Speech acts*. New York: Academic Press.

Grodner, D. J., N. M. Klein, K. M. Carbary, and M. K. Tanenhaus (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition 116*(1), 42–55.

Horn, L. (1972). *On the Semantic Properties of Logical Operators in English*. Ph. D. thesis, UCLA.

Huang, Y., E. Spelke, and J. Snedeker (2013). What exactly do number words mean? *Language Learning and Development 9*(2), 105–129.

Huang, Y. T. and J. Snedeker (2009). Online interpretation of scalar quantifiers: Insight into the semantics-pragmatics interface. *Cognitive psychology 58*(3), 376–415.

Katsos, N. and D. V. Bishop (2011). Pragmatic tolerance: implications for the acquisition of informativeness and implicature. *Cognition 120*(1), 67–81.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistic and Philosophy 30*(6), 669–690.

Kennedy, L., C. Bill, F. Schwarz, R. Folli, S. Crain, and J. Romoli (to appear). Scalar implicatures vs presuppositions: the view from Broca's aphasia. In *Proceedings of NELS 45*.

Magri, G. (2010). *A theory of individual-level predicates based on blind mandatory scalar implicatures*. Ph. D. thesis, Massachusetts Institute of Technology.

Romoli, J. and F. Schwarz (2015). An experimental comparison between presuppositions and indirect scalar implicatures. In F. Schwarz (Ed.), *Experimental Perspectives on Presuppositions*, Studies in Theoretical Psycholinguistics, pp. 215–240. Dordrecht: Springer.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy 27*(3), 367–391.

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

Schlenker, P. (2012). The semantics/pragmatics interface. unpublished manuscript ENS, NYU.